

# A chorus of disapproval

The fight against AIDS is losing ground, but the current spate of mud-slinging is far from helpful.

**T**he global conversation about AIDS is beginning to sound like a high-decibel exercise in finger-pointing and blame. This dangerous trend should be on the minds of the thousands of attendees convening in Mexico City this weekend for the XVII International AIDS Conference.

Thirty-three million people around the world are HIV-positive, and more than 6,800 become infected every day. Tests on microbicides and vaccines have failed, and have put some volunteers at greater risk of HIV infection. Yet critics are attacking the very programmes and people trying to solve these problems, with some even calling for an end to government spending on the search for a vaccine.

This is an overreaction. As many scientists point out, the search for a malaria vaccine has seen dozens of failed trials, whereas only three AIDS vaccines have so far been tested in efficacy studies. What is needed are better vaccine candidates to test, so it makes sense that the major backers of HIV vaccine trials, including the US National Institutes of Health, are now focusing on the basic research that could help the field move forward (see page 565).

Meanwhile, two books published last year claim that the United Nations AIDS programme, UNAIDS, has led an ineffective, politically motivated response to the disease and has distorted statistics in an effort to garner more money (see *Nature* 447, 531–532; 2007). And critics such as Roger England, who runs a small think tank in Grenada, argue that spending on AIDS has distorted poor countries' priorities and weakened their health systems. England proposes that UNAIDS be shut down, and the money spent on AIDS programmes shifted to general funding for health systems. Amid the debate on these questions, the founding director of UNAIDS, Peter Piot, announced in April that he would step down at the end of this year, throwing the agency into uncertainty at a crucial time.

There is no doubt that many poor countries' health systems are struggling, but it is wrong to say that AIDS aid is responsible. In fact, AIDS programmes have shown how poor countries can use new models to deliver needed care, for instance by providing antiretroviral treatments effectively, putting to rest claims that the costly drugs could not be used correctly outside resource-rich nations.

It is also wrong to assume that governments will spend money effectively to fight AIDS if given funds to support health systems overall, as England suggests. Today, many strategies for delivering AIDS treatment target groups such as women, homosexuals and intravenous drug users that have been ignored by governments in the past — neglect that fuelled the spread of the disease. More money should be spent on both AIDS and strengthening health-care systems. And this will be possible if donor governments live up to their promises, such as the pledges of general and disease-specific aid to Africa that were repeated this July at the G8 meeting in Japan.

On that front, it is heartening that the US House and Senate have reauthorized \$48 billion for the President's Emergency Plan for AIDS Relief (\$9 billion of which is for fighting malaria and tuberculosis). If President Bush signs the bill as expected, the programme will also permit the US government to reverse the shameful and embarrassing policy that bans travellers with HIV from entering the country. That might serve as an example to other governments that still sanction discrimination against those who are HIV-positive.

The world is still far from achieving the goal adopted in 2000 by UN member states, which pledged to provide universal access to AIDS treatment by 2010. Three million people now receive life-saving antiretroviral drugs, but 70% of those in low- to middle-income countries who need them don't get them. Indeed, the example of wealthy nations themselves shows what happens when they lose focus on AIDS. In the United States, for instance, reports now indicate that HIV infection rates have begun to rise in Latinos and young gay men.

The activists and scientists about to meet in Mexico City must demand that leaders keep their eye on the ball. The world now has models for providing treatment and care in the places that sorely need it, and is in a position to make more tangible gains against AIDS. This is no time to backslide, and the Mexico City meeting must deliver this message loud and clear. ■

**"More money should be spent on both AIDS and strengthening health-care systems."**

## The greener grid

Governments need to back an overhaul to get the electricity grid ready for renewable energy.

**O**n 17 July, former US vice-president Al Gore proposed that the United States should commit to producing 100% of its electricity from renewable energy and truly clean carbon-free sources within ten years. That is an ambitious and optimistic plan, to put it mildly. But whether or not the United States

achieves Gore's goal, the country is already expanding its generation portfolio to include renewable sources of electricity, such as solar, wind and geothermal, and could do so much faster with a better regulatory and incentive system. Successful examples of the latter can be found in Germany, where 'feed-in tariffs' offer end-users a guaranteed price for selling renewable electricity back to the grid (see page 558).

True, these shifts towards greener sources of electricity can have side effects that are not so green. A case in point is Texas, whose famous winds are now producing more than 5,300 megawatts, by far the largest installed wind-turbine capacity in the United States.

Another 2,000 megawatts of capacity are under construction — and the pace may well pick up. Billionaire oilman T. Boone Pickens, who has invested heavily in wind, has recently released a series of television adverts and appeared on news programmes touting 'the Pickens Plan', in which the country would shift towards wind for electricity and natural gas for transportation.

Unfortunately, the strongest and most reliable wind is often found far from customers. That is why the Texas wind farms cluster in the gusty Panhandle region, hundreds of kilometres from the centres of population farther east. So earlier this month, Texas approved a nearly \$5-billion plan to build new transmission lines linking wind farms to customers. Wind-energy proponents say this will greatly ease the bottleneck limiting the development of wind resources in the state, and might even allow Texas's installed wind-power capacity to overtake Germany's — currently the world leader at more than 22,300 megawatts. Yet this triumph for renewable energy will be a bitter pill for many environmentalists, as it will mean large transmission lines cutting through once-natural landscapes.

Wind, like the sun, is also fickle. In February, for example, a sudden lull idled the Panhandle's wind farms and forced managers of the Texas grid to cut power to some large customers for an hour and a half. So another sour note for environmentalists is that renewable power sources may have to be backed up by generators that are more reliable, such as natural gas and cogeneration plants.

Yet these not-so-green side effects are not the end of the story. Giant wind farms are something of an exception in the renewable-energy picture, as they resemble the traditional model of big, centralized coal, gas or nuclear power plants feeding bulk-distribution networks.

Looked at more broadly, renewables point to a future in which the power grid could be far less centralized than it is today, with a much greater reliance on local power sources such as rooftop solar panels. This could make the electricity grid more efficient, as less power would be lost in long-distance transmission.

Such a grid would benefit from better electricity storage — in the form of large-scale batteries, say — to smooth out drops in wind or cloudy days. And it would certainly have to be smarter than it is today. To manage this decentralized proliferation of sources and users, the grid would have to be liberally studded with microprocessors that can take actions on their own, without humans pressing any buttons or picking up any phones (see page 570).

Public and private programmes to achieve these goals will probably generate many new, highly skilled 'clean-tech' and 'green collar' jobs. Both candidates for the US presidency have talked about investing in the grid, and Democrat Barack Obama has specifically promised to spend \$150 billion to create 5 million green jobs if elected. Meanwhile, countries such as Germany, Denmark, Britain and Australia are already showing that green jobs can buoy their economies.

Serious attention to the newer, greener grid should be a continuing priority for governments and private investors around the world. With effective investment, regulations and incentives, the enormous task of remodelling the grid for renewables will be a boon to both the environment and the global economy. ■

**"Renewables point to a future in which the power grid could be far less centralized than it is today."**

## Europe's science forum

Although the Euroscience Open Forum is a meeting to be proud of, its organizers should aim even higher.

When a small group of academics founded the organization 'Euroscience' in 1997, few would have imagined it could acquire the clout and funds to drive a regular biennial meeting, now attracting 5,000 attendees, in a major European city. The grass-roots society (see [www.euroscience.org](http://www.euroscience.org)) risked being dismissed as a small band of Euro-wonks, although some organizations (including this journal) welcomed its arrival. But last week's third Euroscience Open Forum (ESOF, [www.esof2008.org](http://www.esof2008.org)) in Barcelona, Spain, showed that Euroscience's sheer determination in the face of such perceptions has again paid off. Anyone attending the meeting will have been struck by the energy on display, the high attendance of young people, and the active engagement with the media.

*Nature* was involved in the meeting in several ways, and cannot claim detachment. Nevertheless, informal canvassing of opinions of other participants reinforced the notion that this ESOF meeting was both lively and — to judge by the European movers and shakers present — important. The attendance of participants from the United States and elsewhere outside Europe was also a good sign.

It is still an open question just how influential this forum might

become, rather than simply acting as a showcase. That question requires the attention of the organizers of the next ESOF, to be held in Turin, Italy, in 2010. The science sessions in Barcelona tended to communicate what is happening, but not to address challenging issues. There could have been more heavyweight political figures in attendance, but there was a large turnout by representatives from funding bodies, the European Commission, and national and European parliamentarians. There were also substantive policy discussions. These included the first opportunity for European scientists and others to respond critically and constructively to the achievements and plans of the European Research Council, and also to discuss forward-looking scientific thinking of the European Science Foundation.

Critics pointed to an under-representation of social sciences and of eastern countries, and also the usual disparity in impact between strong sessions where the moderators had clearly made an effort to work coherently with presenters beforehand, and lamentable sessions where they hadn't. The lack of plenary women speakers was much criticized, although insiders said that they had had a high refusal rate to invitations.

The fact that ESOF is seen as an important opportunity by host cities is evidenced by the fierce competition already under way to host the meeting in 2012 — one couldn't move without being confronted by a Dubliner, it seemed, although the Viennese were vociferous too. ESOF can and should build on such ambitions to become even more significant for Europe's citizens — not least its scientists. ■

# RESEARCH HIGHLIGHTS

## Storm warning

*Science* doi:10.1126/science.1160495 (2008)

A realignment of Earth's magnetic field lines rather than a disruption of current across its magnetosphere triggers the geomagnetic substorms that lead to luminous polar auroras. That, at least, is what the first results from NASA's five-satellite THEMIS mission suggest.

The magnetosphere is a volume enveloping the planet that protects it from a stream of charged particles from the Sun. Before a substorm, Earth's magnetic field becomes entwined with this solar wind. When the field snaps back into place, it prompts the chain of events that causes the substorm.

Vassilis Angelopoulos of the University of California, Los Angeles, and his team used THEMIS to observe a nascent substorm in February. They found that the snapping back, or realignment, of the magnetic field preceded a brightening of the aurora. Disruption to the magnetospheric current came after that.



C. ANDERSON/AURORA CREATIVE/GETTY

## MOLECULAR BIOLOGY

### Telling time

*Cell* **134**, 317–328 (2008); *Cell* **134**, 329–340 (2008)

A protein that is associated with metabolism and short lifespan also regulates the body's internal clock, two research groups report.

Ueli Schibler at the University of Geneva in Switzerland and his colleagues discovered that the protein, called SIRT1, is needed for normal expression of several important clock genes. Meanwhile, Paolo Sassone-Corsi of the University of California, Irvine, and his colleagues found that deleting the gene that encodes SIRT1 from the DNA of liver cells disrupted the circadian rhythms of mice.

Both teams also showed that SIRT1 forms a complex with CLOCK, a protein that, as its name suggests, has a key role in regulating the body clock. Because SIRT1 activity is dependent on NAD<sup>+</sup>, a cellular metabolite, the protein provides a molecular link between metabolism and the circadian clock.

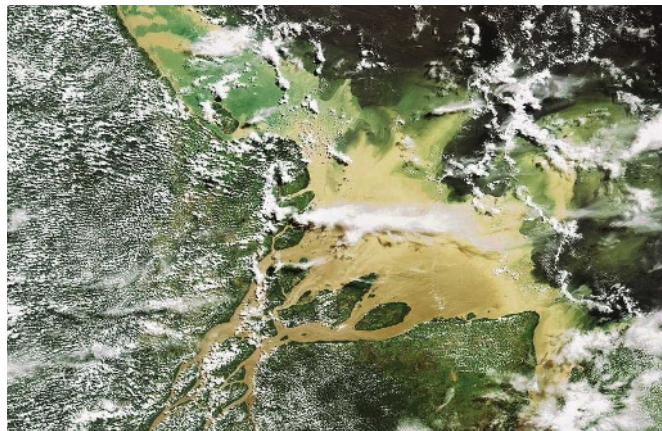
## ENVIRONMENTAL SCIENCES

### Fresh data

*Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0710279105 (2008)

Murky plumes of Amazonian fresh water are carried hundreds of kilometres out to sea (as pictured by satellite, right). Whether these waters suppress or enhance primary production in the Atlantic has been up for debate.

With this in mind, and with data from 82 oceanic field stations to hand, Ajit Subramaniam of Columbia University in New York and his co-workers have labelled



the Amazon's discharge a carbon sink. They calculate that photosynthetic nitrogen-fixing organisms that assimilate iron and phosphorus in the plume — and then die and fall to the sea floor — sequester 20.4 million tonnes of carbon per year. A further 7.2 million tonnes of carbon is fixed annually by organisms that use nitrate delivered to the ocean by the river.

## PHYSICAL CHEMISTRY

### Over ice

*Phys. Rev. Lett.* **101**, 036101 (2008)

Might water freeze spontaneously in tight spaces? K. B. Jinesh and Joost Frenken at Leiden University in the Netherlands say that a film of water just a few molecules thick turns to ice at room temperature when it is confined between a piece of graphite and a blunt tungsten needle.

They write that the needle tip moved jerkily across the graphite, similarly to chalk across a blackboard, rather than being lubricated by the intervening water. The jerks suggest a surface corrugation closer to that of ice than

graphite. Given previous controversy about how water behaves in nanoscale spaces, the claim is likely to excite lively debate.

## NANOTECHNOLOGY

### Weighing options

*Nature Nanotech.* doi:10.1038/nnano.2008.200 (2008)

Physicists at the University of California, Berkeley, have developed a way to make carbon nanotubes weigh things. Like all objects, nanotubes vibrate with their own characteristic frequency, which is related to their mass. So if a tiny object attaches to the nanotube, making it heavier, the nanotube's resonant frequency will decrease.

Kenneth Jensen and his colleagues set a nanotube in a vacuum and released gold atoms into the chamber. When one of these landed on the nanotube, its vibration rate changed.

The technique is sensitive to frequency changes that correspond to mass increases of 10<sup>-25</sup> kilograms. This is not as accurate as conventional mass spectrometry, but does not require a sample to be ionized, a process that can destroy biological molecules.

## QUANTUM COMPUTING

### Cloudy computing

*Phys. Rev. Lett.* **101**, 040501 (2008)

Quantum computing is in its infancy, in part because creating a large number of quantum bits, or 'qubits', is impossible without better control over quantum states. Side-stepping this snag, Klaus Mølmer and his colleagues



at the University of Aarhus in Denmark propose encoding lots of qubits in clouds of polar molecules.

Their theoretical set-up would use a laser to 'write' qubits as a pattern across an entire molecular cloud — an easier task than controlling individual molecules. Radio waves would then transmit the data to a tiny patch of superconducting material where calculations could be performed. However, aligning the cloud with the radio antenna will be tricky, Mølmer warns.

## MOLECULAR BIOLOGY

### Senior signals

*Cell* **134**, 291–303 (2008)

By comparing the long list of messenger RNAs produced in young and old roundworms, biologists have identified three proteins that control ageing independently of environmental insults and the cellular wear and tear that accumulates over time. The proteins, ELT-3, ELT-5 and ELT-6, participate in the senescence process of *Caenorhabditis elegans* as a part of the animal's own developmental program.

Stuart Kim of Stanford University Medical Center in California and his team silenced the genes encoding ELT-5 and ELT-6 in some of their worms. This showed that these proteins lower ELT-3 expression in adult worms. ELT-3, the researchers found, regulates a host of downstream genes involved in turning transparent and spritely young worms into pigmented and more flaccid oldsters.

## PHOTONICS

### Telescopic TV

*Nature Photon.* doi:10.1038/nphoton.2008.133 (2008)

Telescopes often have a ring-shaped primary mirror to concentrate light onto a secondary mirror, which, in turn, reflects it through the hole in the primary. The same principle informs a new design for backlit screens.

In the 'telescopic pixels' described by Anna Pyayt of the University of Washington in Seattle and her team, the shape of the primary mirrors is under electronic control. When their shape is near-parabolic, they bounce light onto the secondaries, illuminating the holes in the primaries. When they are flat, no light hits the secondaries and the holes remain dark.

The prototype pixels transmit more than three times as much light as comparable liquid crystal displays. At present, the contrast they achieve is low, but models suggest that this can be improved. The technology may be suited to large, energy-efficient flat panels.

## IMMUNOLOGY

### Green vaccines

*Proc. Natl Acad. Sci. USA* **105**, 10131–10136 (2008)

Ronald Levy of Stanford University Medical Center in California and his colleagues have carried out the first clinical trial of plant-produced therapeutic vaccines against non-Hodgkin's lymphoma. The vaccines were grown in tobacco plants by infecting the plants with a tobacco mosaic virus containing DNA fragments cloned from each patient's cancer cells, thus matching each vaccine to its potential beneficiary.

Eleven of the 16 participants with the cancer had developed a cellular immune response 28 days after vaccination; six of these made antibodies specific to certain proteins found on the surface of their cancerous white blood cells.



## ZOOLOGY

### Bigmouth strikes again

*J. Fish Biol.* **73**, 17–34 (2008)

An examination of a one-tonne megamouth shark (*Megachasma pelagios*; pictured above) caught off the coast of Japan indicates that the species has a feeding method previously seen in whales, but never before in sharks.

Fellow plankton-feeding sharks — the whale shark and the basking shark — typically feed by swimming through clouds of their quarry with their mouths continuously open. But the highly elastic skin and loose connective tissue around the jaws of the megamouth point to it relying on an 'engulfment' technique like that used by humpback whales, say Kazuhiro Nakaya and his colleagues at Hokkaido University in Japan.

This would involve the fish swimming towards a shoal of plankton with its mouth slightly open, sucking, and then gradually expanding the mouth and throat cavity to fit more and more in. When its mouth is full, the shark 'gulps' — sieving and swallowing prey while expelling seawater through its gills.

T. YAMANAKA/AFP/GETTY

## JOURNAL CLUB

Robin Rogers

The University of Alabama, Tuscaloosa; Queen's University Belfast, Northern Ireland

**A chemist believes that an ionic liquid is the place for a noxious gas.**

As a 'green chemist', I worry about the potential dangers of moving toxic and flammable gases around. Most nasty gases are transported in pressurized canisters to save space, posing the risk of hazardous compounds being expressed over people and pleasant greenery on the rare occasions that a container breaks.

Recently, some scientists at Air Products and Chemicals, a chemicals supplier in Allentown, Pennsylvania, found a way to store phosphine ( $\text{PH}_3$ ) and boron trifluoride ( $\text{BF}_3$ ) — both toxic gases — in ionic liquids, and then recover the gases without introducing impurities (D. J. Tempel *et al.* *J. Am. Chem. Soc.* **130**, 400–401; 2008).

The advantage of transporting gases in ionic liquids is that many such liquids have no measurable vapour pressure. So were a container to burst, the gases inside it would remain as chemical complexes in a liquid state, making them much easier to mop up. Furthermore, ionic liquids can be recycled in subsequent shipments.

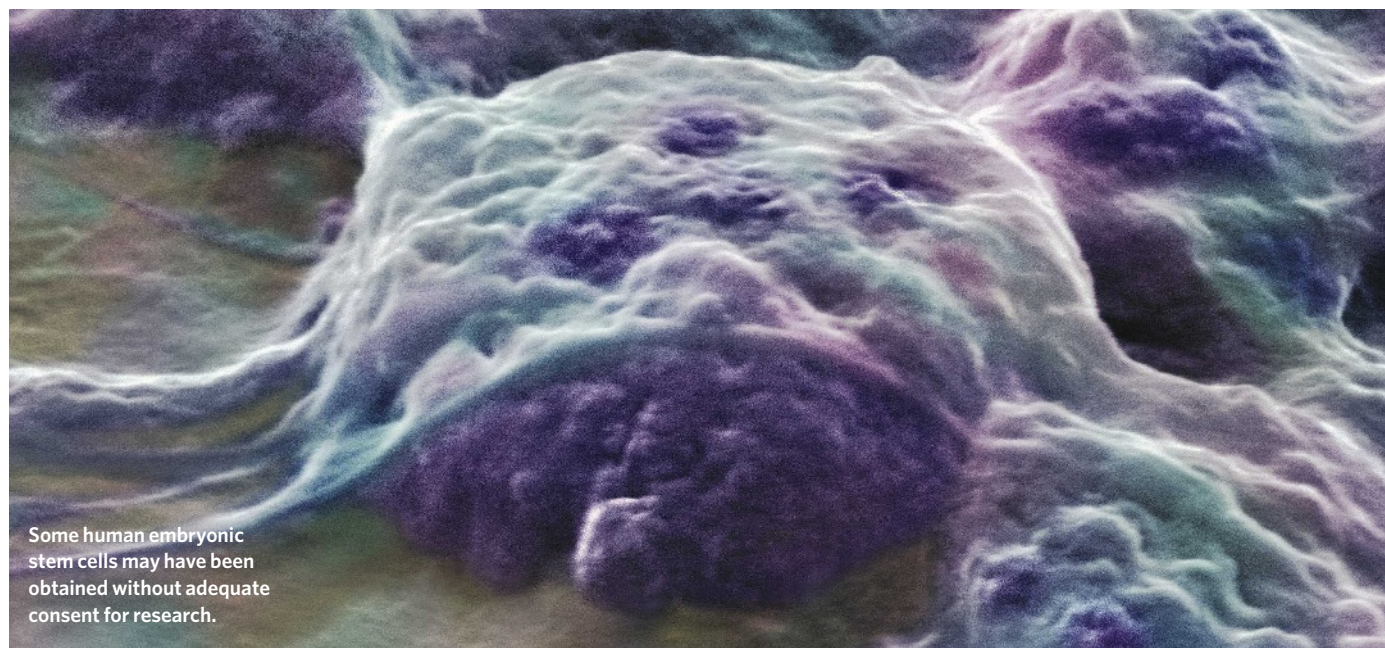
Dan Tempel and his team used a computer model to consider two ionic liquids — the cation 1-butyl-3-methylimidazolium paired with either  $\text{Al}_2\text{Cl}_7^-$  and  $\text{Cu}_2\text{Cl}_3^-$  — for phosphine transport. They then tested the latter in the lab; the positively charged copper atoms bound the lone electron pair on phosphine. Similarly, the electron-deficient boron atom in boron trifluoride facilitated the formation of a covalent bond with a fluorine atom in another ionic liquid, in which the same cation is paired with  $\text{BF}_4^-$ .

In both cases, more than 90% of the ionic liquid's reactive sites formed complexes at room temperatures. This means that relatively small volumes of ionic liquids could move a lot of toxic gas around. I think this could revolutionize the industry.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



## NEWS



A. CAVANAGH/WELLCOME IMAGES

Some human embryonic stem cells may have been obtained without adequate consent for research.

## Consent issues restrict stem-cell use

Stanford University is to tell its researchers that around one-quarter of the human embryonic stem-cell lines eligible for US government funding are now off-limits because of ethics concerns.

The institute, in Palo Alto, California, is concerned that some of the women who donated the embryos for these stem-cell lines did not give informed consent for the lines to be used in research. Johns Hopkins University in Baltimore, Maryland, has decided to reconsider lines individually as researchers express an interest in using them.

The concerns follow an analysis by bioethicist Robert Streiffer of the University of Wisconsin, Madison, who obtained copies of the informed consent forms given to donors of the 21 lines that have been approved for federal research funding by the US National Institutes of Health (NIH). Streiffer compared these forms with guidelines set by the US National Academy of Sciences (NAS). None of the forms met the guidelines exactly, he concluded, and some deviated egregiously (R. Streiffer *Hastings Cent. Rep.* **38**, 40–47; 2008).

Now, ethics oversight committees at universities across the United States are questioning which lines should be permissible for research — and hoping that another agency, such as the NIH or a state government, will make the decision for them.

On the basis of Streiffer's analysis, Stanford has decided that as many as five of the lines should not be used, it has emerged. The lines

were derived by two biotechnology companies — BresaGen, based in Athens, Georgia, and Cellartis in Göteborg, Sweden — both of which say they obtained adequate consent according to standards in place at the time. But rather than signing a separate form authorizing research, the form signed by donors for lines derived by BresaGen (now owned by Novacell) was for fertility treatment and noted that multiply fertilized eggs or embryos that were not developing could be used for scientific study. The consent form used by Cellartis stated that cells would be destroyed after a few days in culture. Neither consent form stated explicitly that the research would potentially destroy viable embryos.

All the NIH-approved lines were derived before August 2001, when President George W. Bush declared that only lines already in existence could receive federal support. In 2005, responding to calls from the scientific community, an advisory committee of the NAS issued guidelines that covered what kinds of experiments should be conducted on human embryonic stem cells and how donors of gametes and embryos should be treated. Later, the committee declared that the lines on the NIH registry had been derived under conditions “substantially similar” to its guidelines and instructed research oversight committees to consider the lines as acceptably derived.

But members of the committee relied on the NIH's 2001 assessment that the lines were

appropriate for research use. “It didn't occur to us that we should get the consent forms to look at them,” says committee study director Frances Sharples, who says that the NAS will revisit the issue of which lines are acceptable at a meeting later this year.

“We understood that the lines were in substantial compliance with NAS guidelines,” says NAS committee co-chair Jonathan Moreno, who is a senior fellow at the Center for American Progress, a think tank based in Washington DC. “It shows the lengths to which the administration pushed the NIH to get as many lines on

record as it could. It shows that it's time to move on.”

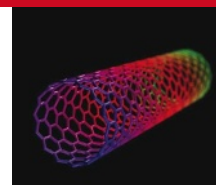
The NIH will not be taking any lines off its registry, says Story Landis, head of the agency's Stem Cell Task Force. “Streiffer's paper deals with application of

2008 standards to cell lines that were put on the registry in 2001,” she says, adding that NIH staff determined that the registry lines met the criteria put forth by the president in 2001.

It's unclear the extent to which reassessment will affect research on human embryonic stem cells. Many scientists with private or state funding are already using newer non-NIH lines derived under better conditions. Of the lines on the NIH registry, the ones from BresaGen and Cellartis are the least used. Nonetheless, they have been shipped to dozens of researchers. ■

**Monya Baker**

For more, see <http://tinyurl.com/6ae7ne>.



**NANOTECHNOLOGY**  
EPA's voluntary safety scheme is undersubscribed.  
[www.nature.com/news](http://www.nature.com/news)

S. BRANDT/ALAMY

# Italy picks businessman to head space agency

Italian scientists are worried that a shake-up of the nation's space agency will put commercial and defence interests ahead of research.

Prime Minister Silvio Berlusconi is in the process of replacing the president of the Italian Space Agency. He is removing the agency's current head, astrophysicist Giovanni Bignami, and installing business executive Enrico Saggese, who heads the space division at Finmeccanica, Italy's largest aerospace firm.



**Giovanni Bignami is head of the Italian Space Agency.**

The move has met with concern from the agency's independent scientific and technical council, most of whom signed a letter to the government last week asking it to reconsider. "We are worried that the scientific programme will be under-funded or cancelled in the next couple of years," says Guido Visconti, an atmospheric physicist at the University of L'Aquila and a member of the council.

The shake-up began in early July when six of seven members of the space agency's governing council resigned to allow the Berlusconi government "the greatest level of freedom" in

determining the agency's future. Under its own rules, the mass resignation made the agency ungovernable and Italy's council of ministers is now seeking to install Saggese as a commissioner — an interim position that would give him complete control. A similar set of circumstances surrounded the resignation of the agency's previous president, aerospace engineer Sergio Vetralla, in 2006.

Berlusconi came to office in May, and the latest political manoeuvring is seen by some as an attempt to strengthen the space agency's ties with industry. Others think that the government may be moving the agency towards providing more military surveillance and other applications. The government did not respond to *Nature's* requests for comment, but has previously denied that it is planning such changes.

Bignami has only been the incumbent for a year, but has been credited with raising the profile of research, which takes up just 10–15% of the agency's roughly €700 million (US\$1.1 billion) annual budget. Coming from a research position at the Institute for Advanced Study in

Pavia, he is widely seen as more science-savvy than his predecessors. And he has given scientists more control of how research funds are spent by instituting a series of review panels. "We think he did a great job," says Pietro Ubertini, director of the Institute of Space Astrophysics and Cosmology in Bologna. "He gave us full freedom to choose the scientific programme."

Bignami is also popular with Italy's foreign partners as, for example, he helped to coordinate data-sharing between international weather satellites. Italy will chair the next triennial European Space Agency ministerial meeting in November, and Bignami's replacement "could be very disruptive to overall planning", warns Bo Andersen, director general of the Norwegian Space Centre.

However, not everyone is concerned. "As far as science is concerned, I don't think it's a big issue," says Guido Chincarini, an astrophysicist at the University of Milan, who points out that the proposed deputy commissioner, Piero Benvenuti, is an astronomer at the University of Padova. He is also one of the resigning directors.

The government has less than a month to install Saggese, but for now, Bignami continues to run the agency. "No one has asked me to resign," he says. "And therefore I have not."

**Geoff Brumfiel, additional reporting by Faye Fornasier**

# School chemistry off-limits to terrorism suspect

A British judge has ruled that a terrorism suspect cannot take secondary-school level courses in chemistry and human biology.

In a first-of-a-kind ruling, High Court judge Stephen Silber said that the courses would put the suspect — who is referred to as A.E. for privacy reasons — in a "substantially stronger position" to carry out chemical and biological attacks.

But scientists are unconvinced. An AS level in human biology is unlikely to provide much in the way of skills to a potential terrorist, says Charles Penn, a molecular microbiologist at the University of Birmingham. "I'm pretty sceptical that this is a real, tangible risk," he says.

The ruling paints a "misleading image of school chemistry being a subject of particular value to

potential terrorists", says Richard Pike, chief executive of the Royal Society of Chemistry. "There is nothing on the AS-level chemistry course that cannot be found easily on the web and through other means," he says.

A.E. is an Iraqi national with alleged ties to terrorist groups including Al-Qaeda in Iraq, according to the UK government. Since 2006 he has been subject to a 'control order', a special legal instrument that places limits on his freedoms, including his Internet use. He is subject to surveillance and must seek approval from the Home Office before undertaking coursework.

In September 2007, he asked for permission to enrol in AS-level courses in chemistry and human biology. In Iraq, A.E. had been

training as a doctor and the courses were the first step in continuing his education, according to his solicitor, Mohammed Ayub, of Chambers Solicitors in Bradford. The Home Office denied his request, claiming that knowledge gained in the coursework could be used for terrorism. A.E. contested the decision in court, claiming that AS-level courses were largely harmless (see *Nature* 450, 467; 2007).

The 15-page ruling, released on 21 July, upholds the Home Office's claim. Based in part on testimony from an anonymous security official known only as 'X', Justice Silber found that A.E. would gain expertise, particularly in using equipment, that might further terrorist activities. But Penn questions how useful the courses would be for a would-be terrorist.

Even university-level students would have trouble producing large quantities of a pathogen such as anthrax, he says, "they wouldn't have a clue where to start".

Silber also called into question A.E.'s previous medical training in Iraq, which had been a key part of A.E.'s argument — his lawyer maintained that the AS-level courses would review information he had already learned. Even if A.E. could learn new techniques from the courses, Ayub adds, he is under such tight scrutiny that he would be unable to do much. "The risk is negligible to nil," he says.

Ayub says that his client is "dismayed" by the ruling, and they are likely to decide whether to take the case to the Court of Appeal within the coming weeks.

**Geoff Brumfiel**



## SPECIAL REPORT

## Thin films: ready for their close-up?

New types of solar cell that can be mass-produced cheaply, and integrated into building materials, are popular with venture capitalists and market analysts. But scientists are less gung ho, reports **Declan Butler**.

From the 1950s onwards, big chunks of crystalline silicon have dominated the world of solar cells. But the dominance of these traditional cells — which make up 90% of today's 10-gigawatt-a-year installation market — is now being challenged by 'thin-film' solar cells that are micrometres or mere nanometres thick, and frequently made of materials other than silicon. Some argue that such a change in technology is the only way that solar-cell technology can hope to maintain the 50% annual growth it has enjoyed during the past five years.

Highly-purified silicon is expensive, and the fact that its other users, silicon-chip makers, manufacture high-value products from it has helped to keep it so. That presents a problem for makers of a commodity such as solar cells, a problem exacerbated by manufacturing processes that waste a significant amount of raw material. Thin films, by comparison, can in principle be made cheaply by using reel-to-reel processes similar to those of a printing press and other mass-production techniques.

They can also be applied on to flexible sheets of steel and other materials, and so be integrated directly into building materials. These advantages, though, come at a cost. Thin films are normally significantly less efficient than traditional silicon cells, which have conversion efficiencies of 15% or upwards in everyday use.

Nevertheless, Nanomarkets, a consultancy based in Glen Allen, Virginia, issued a report on 21 July that predicted thin films taking as much as half of the world market for photovoltaic cells by 2015. Lux Research, a New-York based analyst firm, reports that investment in the sector climbed from US\$481 million in 2006 to \$1.36 billion in 2007.

Most thin-film cells sold today still use silicon, but in its amorphous, rather than crystalline, form. This makes the cells thin and cheap but costs them half or more of their efficiency compared with traditional designs. The hope, and to some extent the hype, is focused on new technologies. Of these, cadmium telluride-based systems are the most mature, and hold one third of the thin-film market, according to

IDTechEx, a consultancy based in Cambridge, Massachusetts. The technology is dominated by First Solar, a company in Phoenix, Arizona. But the venture capitalists' favourite is a technology known as CIGS (copper indium gallium diselenide).

CIGS's semiconductor cells are the province of established companies such as Würth Solar in Germany as well as Showa Shell Sekiyu and Honda in Japan, and the focus of specialized start-ups Miasolé and Nanosolar, in California's Silicon Valley, and Heliovolt in Austin, Texas. The cells are held to marry the advantages of thin-film manufacture with efficiencies close to those of today's best products. In March, a thin-film solar cell developed at the US National Renewable Energy Laboratory (NREL) in Golden, Colorado, set a CIGS efficiency record of 19.9%, although most commercial thin-film CIGS contenders at the moment offer something much closer to 10%.

Nanosolar of Palo Alto, California, perhaps the most bullish of the thin-film CIGS companies, claims to have developed a production

## PHOTOVOLTAIC TECHNOLOGIES

| Solar cell  | Advantages  | Disadvantages   | Status  | Efficiency*        | Type  |
|---|---|---|---|--------------------|---|
| Crystalline silicon   | High efficiencies.  | Wafer-based cells use large amounts of expensive silicon, can't be applied as thick film. | Mature technology, long established at industrial scale.  | 15–22%             | Semiconductor. Photons of sunlight are absorbed by a semiconductor, their energy is converted to electrons and positive 'holes' that produce the photocurrent in solar cells, and a photovoltage — both are needed for power. |
| Multi-junction gallium indium phosphide (GaInP), gallium arsenide (GaAs) and germanium (Ge) | Traditionally used in space, as has highest efficiencies.   | Expensive crystalline layer production.   | At research stage; limited application, but trend towards increased terrestrial applications in combination with light concentrators. | 30–40%             | Semiconductor.  |
| <b>Thin film</b>  | Conventional crystalline silicon cells use solid wafers a couple of hundred micrometres across. Thin-film technologies, which can use inorganic semiconductors, organic dyes or organic polymers, are nanometres to a few micrometres thick, and can be applied in continuous industrial production using, for example, vacuum deposition, sputtering and printing. Thin films also use fewer raw materials compared with silicon wafers, and so cut costs. |   |   |                    |   |
| Nano and amorphous silicon  | Can be used as thin film unlike crystalline silicon.  | Lower efficiency.   | Uncertainties over long-term stability.   | 6–10%              | Semiconductor.  |
| Cadmium telluride (CdTe)  | Reasonable efficiencies and proven technology. Good at low light levels.  | Need careful disposal of toxic cadmium.   | Mature, large applications started.   | 9–11%              | Semiconductor.  |
| Copper indium gallium diselenide (CIGS)   | Highest efficiency of thin films, can be made transparent. Good optical and electronic properties.  | Production processes still relatively untested.   | Moving to industrial scale.   | 10–14%             | Semiconductor.  |
| Dye sensitized solar cells (DSSC) — 'Grätzel cell'  | Work well at low light levels, and extreme angles. Transparent.   | Solvent electrolyte handling.   | Moving from development into deployment.  | 11% (Grätzel cell) | Not a semiconductor. Uses processes mimicking photosynthesis (see text).  |
| Organic   | Potentially the cheapest solar cell, allowing incorporation in buildings, clothes and devices. Can be transparent.  | Generally poor efficiency. Poor stability.  | Research stage.   | Typically 2–5%     | Uses organic semiconductors, such as polymer fullerenes.  |

\*Efficiencies given are industry estimates. Table compiled from personal communications and from reports by IDTechEx.com, luxresearchinc.com and nanomarkets.net.





## HYDROGENASE REVEALED

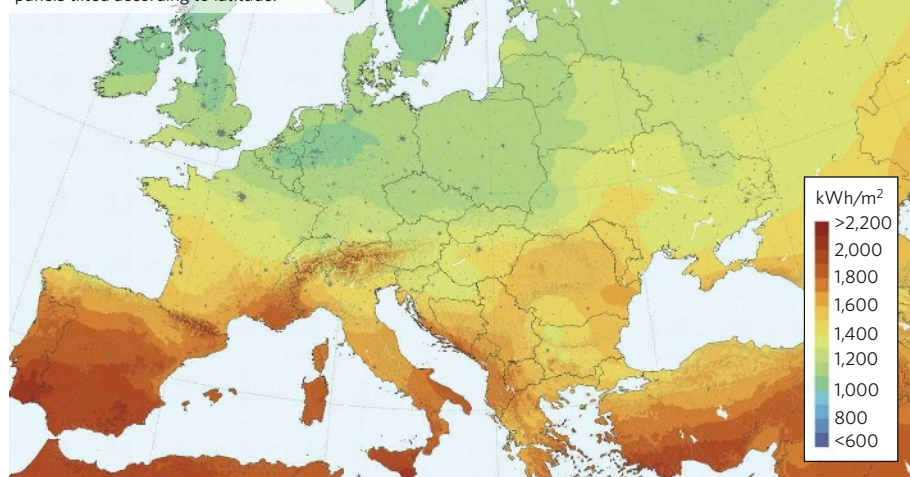
Enzyme structure reveals key ingredients for making hydrogen.

[www.nature.com/news](http://www.nature.com/news)

NOAA

## SOLAR POSSIBILITIES

Annual solar radiation energy in Europe for panels tilted according to latitude.



PVGIS/SÚRI, M. ET AL. SOL. ENERGY 81, 1295–1305, 2007.

system that, for a capital cost of \$1.65 million, could churn out 600 metres of solar panels per minute, for an annual yield of one gigawatt of capacity. This is orders of magnitude cheaper and faster than conventional thin-film vacuum deposition techniques: *Forbes* magazine has reported that Showa Shell Sekiyu, part of Shell's Japanese arm, is looking at spending 100 billion yen (\$930 million) on a one-gigawatt-a-year solar-cell facility using more traditional thin-film manufacturing.

### In development

Some scientists are sceptical as to whether CIGS will, in practice, live up to its extravagant promise. Mass-producing CIGS films with consistent microstructure and composition is a challenge that in effect may trade off the benefits of increased manufacturability and increased efficiency. "CIGS is a very difficult technology to manufacture, having larger moisture sensitivity compared with other technologies," says Martin Green, a solar-cell researcher at the University of New South Wales, Australia. Printing of solar cells, *à la* Nanosolar, "is still at the R&D stage", says Tonio Buonassisi, a photovoltaics researcher at the Massachusetts Institute of Technology in Cambridge. "In industrial production today, expensive vacuum equipment is used, for example for plasma-enhanced chemical vapour deposition or sputtering," he says. Buonassisi is also sceptical about the cited high efficiencies for CIGS, arguing that in mass production they so far come in at around just 9%.

Another frequently voiced concern is that there may be material constraints on the technology. Highly refined silicon may be pricey, but in its unprocessed form it is the second most abundant element in Earth's crust.

Large-scale deployment of CIGS could lead to shortages and price hikes in its scarcer raw materials, says Buonassisi. Martin Roscheisen, chief executive of Nanosolar, sees such fears as overblown: "The cost of indium is less than 1% of revenue [for Nanosolar]."

Nor are these thin film designs the only alternatives to crystalline silicon. Grätzel cells, named for their inventor Michael Grätzel at the Swiss Federal Institute of Technology in Lausanne, use dyes painted onto the surface of nanometer-size particles of titanium dioxide. These are immersed in an electrolyte to produce current in a way more like that used in photosynthesis than the semiconductor technology of normal photovoltaics. As yet, the technology has proved hard to commercialize, but companies such as Konarka in Lowell, Massachusetts, and G24i, in Cardiff, UK, continue to work on it. And Grätzel now thinks he has overcome some of the problems using a modified cell that uses a salt melt as the electrolyte (Y. Bai *et al. Nature Mat.* 7, 626–630; 2008).

Still, Buonassisi is one of several scientists who argue that both economics and new research are on the incumbent technology's side. Vast new silicon refineries in China and elsewhere, expected to be online by the end of the decade, promise a glut of silicon. That will reduce the cost of conventional silicon solar cells. And new research is also opening up possibilities of combining the best of both worlds: making thin films with crystalline silicon and thus benefiting from its attendant higher efficiencies without all the cost. Stefan Glunz and colleagues at the Fraunhofer Institute for Solar Energy Systems in Freiburg, Germany, Europe's largest solar research institute, are working on a process that melts a fine layer of high-quality silicon

on to a substrate of lower grade silicon or ceramic. Prototypes of these cells, which can be mass produced in the same way as some CIGS thin films, have this year yielded solar cells 40 micrometres thick with an efficiency of 20%. Innovalight, a startup in Sunnyvale, California, is developing silicon cells that make use of nanoparticle dyes that could also result in high-performance silicon thin-films produced in high-throughput processes.

"The thin film era will also be about silicon," says Daniel Lincot, a solar researcher at the National Higher School of Chemistry in Paris. "We are set to enter the post-bulk-silicon era, rather than the post-silicon era."

### Reducing costs

More broadly, academic researchers consider the current thin films as a second generation of photovoltaics, but not the end-point of all development. "They lower the cost per unit area but do not reduce the cost per kilowatt hour that much because the efficiencies are generally lower or not increased over silicon," says Arthur Nozik, a photovoltaics expert at NREL. Solar electricity currently costs around 30 cents per kilowatt hour, and thin films will reduce this to 15–20 cents per kilowatt hour, but that's still higher than the 5 cents per kilowatt hour of electricity produced from coal.

Happily for the industry, though, solar cells installed at home or for a specific business do not compete with the cost of producing electricity, but with the price at which utilities subsequently sell it. In sunny climes, 'grid parity' — electricity generated by photovoltaics as cheaply as it is sold by utilities — is expected within four years or so. And subsidies for solar energy have proved politically acceptable in some places, at least so far. In Germany, generous feed-in tariffs that guarantee a high price for solar energy sold back to the grid have encouraged the technology's uptake and have stimulated an indigenous industry in cell and panel making and installation. But the subsidies involved are expensive, and were trimmed earlier this year. A future in which such feed-in tariffs taper off as solar cells get cheaper to install could wean the sector from subsidies without causing a collapse.

"There is a place and market niches for every one of these technologies. There won't necessarily be only one winner," says Harry Zervos, a photovoltaics technology analyst with IDTechEx. That does not mean there will not be losers. Analysts expect a shake-out in the field — through failures and mergers — to start soon. But those who come through should be well placed: "If solar energy doesn't lift off at this point in time," says Zervos, "it never will."

See Editorial, page 551.



NASA AT 50

Past glories and future challenges.

[www.nature.com/news](http://www.nature.com/news)

BRIDGEMAN ART LIBRARY/GETTY IMAGES

Could the calendar for the ancient Olympic Games have been set by the Antikythera Mechanism?

# Complex clock combines calendars

The Antikythera Mechanism, a clockwork device made in Greece around 150–100 BC, astounded the world two years ago when scientists deduced how this machine was used to make complex astronomical time-reckonings. Now they say that the instrument, discovered in 1901 in a Mediterranean shipwreck, did much more than that.

Researchers have been trying to decode the mechanism's inscriptions and functions for several years. Their latest findings reveal that it links the technical calendars used by astronomers to the everyday calendars that regulated ancient Greek society — most strikingly, the calendar that set the timing of the Olympic Games.

"The mechanism is full of surprises," says Alexander Jones of the Institute for the Study of the Ancient World in New York, who is one of the decoding team. "The latest revelations establish its cultural origin for the first time."

The Olympic Games marked the beginning of a four-year timespan called an Olympiad: a calendar system shared by all the Greek city-states, bringing some uniformity to the chronology of the Hellenistic world. The Games began on the full Moon closest to the summer solstice, which meant that calculating the timing required expertise in astronomy. The Olympic Games were the most prestigious of four sets of games in ancient Greece, called the Panhellenic Games. First recorded in 776 BC, the ancient Olympic Games continued until they were extinguished by a ban from the Christian Roman emperor Theodosius I in around 394 AD.

The latest decoding of the Antikythera Mechanism (see page 614), by British mathematician Tony Freeth of the film company Images

First in London and his colleagues, casts fresh light on the way these calendar schemes were planned, used and integrated<sup>1</sup>. The device had intermeshed toothed wheels that represent calendar cycles. By turning the wheels, a user could figure out the relationships between astronomical cycles to deduce the relative positions of the Sun and Moon and forecast eclipses.

But after two millennia under the sea off the island of Antikythera, near Crete, all that remains of the device are 82 fragments of flaking bronze, including parts of 30 gear-wheels<sup>2</sup>. The numbers of gear teeth are crucial, but must be inferred from the partial wheels that remain. And most of the inscriptions are hidden under corrosion and surface accretions. To read them, the researchers used a method called microfocus X-ray computed tomography, which provides X-ray images of slices through the sample, revealing inscriptions buried beneath the mechanism's surface.

In 2006, Freeth was part of a team that used this and other techniques to figure out much of the mechanism's function, showing it to be an instrument of unparalleled sophistication in antiquity, more or less unrivalled until the clockwork mechanisms of the later Middle Ages<sup>3</sup>.

Now they say that the device was even more sophisticated than that — it unites abstruse astronomical determinations of time with the calendar of civic society. Another ancient Greek calendar cycle, called the Metonic cycle, was established to cope with the incommensurability of the lunar cycle and the solar year — the period of Earth's rotation around the Sun, as determined, say, by the time between successive summer solstices. One Metonic period

is equal to 235 lunar months, which is almost exactly 19 solar years. The Metonic cycle, thought previously to be used only by astronomers, is represented on a dial on the Antikythera Mechanism. But this dial now turns out to be inscribed with the names of months in a regional calendar used in Corinthian colonies in northwest Greece — providing evidence that the device was used for mundane reckonings, and giving a surprising clue to its origin.

**"The mechanism is full of surprises."**

As most of the cargo of the Antikythera wreck was from the eastern Mediterranean, researchers had thought that this was where the Mechanism originated too. But Freeth and his team now think that the instrument may have come from Syracuse in Sicily, the Corinthian colony where Archimedes devised a planetarium in the third century BC. "Archimedes died at the siege of Syracuse in 212 BC, so we are confident that he did not make the mechanism," says Freeth. "But it is possible that it came from a heritage of instrument-making that originated with him in Syracuse. It is an attractive idea, but purely speculative at present."

The Olympiad cycle appears on a subsidiary dial inside the Metonic dial, previously thought to represent an improvement on the Metonic cycle called the Callippic cycle. "It really surprised us to discover that it showed the four-year cycle of ancient Greek games," says Freeth. "The mechanism always produces more surprises." ■

**Philip Ball**

1. Freeth, T., Jones, A., Steele, J. M. & Bitsakis, Y. *Nature* **454**, 614–617 (2008).
2. Marchant, J. *Nature* **444**, 534–538 (2006).
3. Freeth, T. et al. *Nature* **444**, 587–591 (2006).

See <http://tinyurl.com/6f83dt> for a video.



A. ZITZMANN

## SCORECARD

**Sippin' whiskey**

German scientists revealed 13 previously undiscovered aroma compounds in Kentucky bourbon as well as old favourites such as whiskylactone.

**Sippin' nectar**

The fermented nectar of the bertam palm, as drunk by Malaysia's pen-tailed tree shrew, is about as alcoholic as beer. Yet video evidence shows no sign of inebriation. Perhaps they just like the taste ...

## ON THE RECORD

**“Dr Mitchell is a great American, but we do not share his opinions on this issue.”**

NASA's response to Edgar Mitchell's claim that it is covering up contact with alien civilizations. Mitchell was a member of the Apollo 14 crew.

## NUMBER CRUNCH

**77%** of Americans accepted that there is solid evidence for global warming in January 2007.

**71%** were of that opinion in April 2008.

**13%** was the drop in the acceptance rate among Republicans.

**2%** was the drop among Democrats.

## WORDWATCH

## LOOSE

The Stanford Linear Accelerator Center may need a new name, because Stanford University will not let the government copyright its current one. An elegantly synonymic suggestion from a SLAC website: the Linearly Organised Optimum Science Establishment.

During August, Sidelines will be off on its summer holiday.

Sources: J. Agric. Food Chem., PNAS, Telegraph.co.uk, Pew Res. Center, SLAC

# Thousands of proteins affected by miRNAs

MicroRNAs (miRNAs) subtly influence a vast number of proteins involved in most key biological processes, according to the first large-scale analyses of how these small pieces of noncoding RNA affect proteins (M. Selbach *et al. Nature* doi:10.1038/nature07228; 2008; D. Baek *et al. Nature* doi:10.1038/nature07242; 2008).

“These papers represent a tour de force in miRNA research, utilizing state-of-the-art technology to tackle a very vexing problem in biology, namely the identification of miRNA target genes,” says Frank Slack, a molecular biologist and miRNA researcher at Yale University.

The studies could help answer long-standing questions about how miRNAs work. These tiny snippets of genome, just 21 to 25 nucleotides long, were once thought to be ‘junk’ material because they are not translated into protein. In fact, in animals they control protein levels through two mechanisms: by breaking down the messenger RNA ‘read-out’ from a gene; and by stopping the messenger RNA from being translated into protein. But previous studies had taken only a global look at how miRNAs affect messenger RNA, because of the technical difficulties in examining how miRNAs affect thousands of proteins at once.

Two teams of scientists have now tackled this problem using a state-of-the-art version of mass spectrometry. The technique, called SILAC (stable isotope labelling with amino acids in cell culture), relies on the use of heavy isotopes to label proteins as they are being produced by cells. The researchers used the isotopes to label thousands of proteins being made in cells that had been forced to overexpress or downregulate certain miRNAs. They then compared labelled protein levels in these cells with levels of corresponding proteins in control cells. One of the teams, co-led by Matthias Selbach and Nikolaus Rajewsky of the Max Delbrück Center for Molecular Medicine in Berlin, modified the SILAC method to label and compare proteins being made by treated and control cells, so that their results would not be skewed by long-lived proteins.

These analyses showed that a single miRNA can indeed dampen levels of hundreds of proteins by impeding their translation — not just by breaking down their messenger RNAs. The studies also examined whether the proteins’

levels changed in ways that had been predicted by computer algorithms and found that the algorithms varied in their predictive abilities.

But the most intriguing finding from the studies is that the effects of miRNAs on proteins are usually quite modest, changing their expression levels by less than twofold. That almost led Steven Gygi, a protein chemist at Harvard Medical School in Boston, Massachusetts, to scotch one of the studies after he used a mass spectrometer to examine the first batch of miRNA-boosted cells. “The experimental data looked just like the controls, so we thought it hadn’t worked,” says Gygi.

He told his collaborator, David Bartel, a molecular biologist and Howard Hughes investigator at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, that he wanted to end the project. But Bartel was undeterred: “After working on miRNAs for seven years, we’ve learned that the data are very subtle,” says Bartel.

A deeper statistical analysis proved Bartel right — the effects were there, but they were often small. Both teams confirmed their results by using knocked-out miRNAs in animal cells.

But although most miRNA effects may be small, they can still be powerful, according to the Delbrück group. The team found one miRNA, ‘let-7b’, that can change production of a key protein, called Dicer, by as much as fourfold. Dicer is involved in making all miRNAs and silencing RNAs (siRNAs), which shut down genes through the process of RNA interference. So let-7b directly or indirectly controls the expression of thousands of genes — a finding that wouldn’t have been seen if the team had looked only at Dicer’s messenger RNA, because most of let-7b’s effects on Dicer occur at the translational level, the team found.

The data paint a picture of miRNAs as regulators that fine-tune protein expression in complex and overlapping patterns. That could complicate the picture for drug developers, because most miRNAs have so many different targets. But this is not necessarily a bad thing, as Selbach points out, because some of the most effective drugs hit multiple molecular targets. “I think it’s a very beautiful finding, because it shows that the system is very robust,” Selbach says. ■

**Erika Check Hayden**

**“I think it’s a very beautiful finding, because it shows that the system is very robust.”**



**GOT A NEWS TIP?**

Send any article ideas for Nature's News section to [newstips@nature.com](mailto:newstips@nature.com)

K. CAMPBELL/GETTY

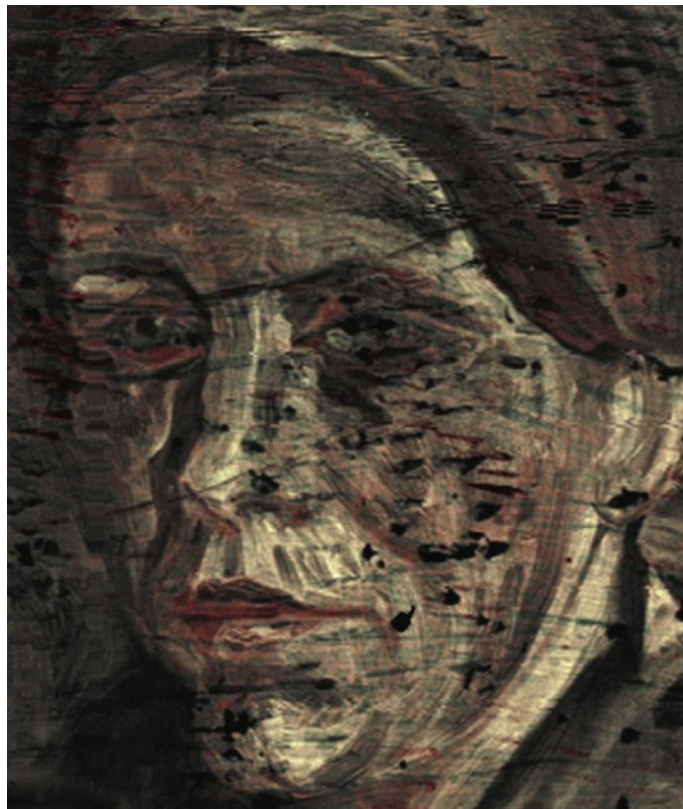
## SNAPSHOT The hidden van Gogh

An unknown Vincent van Gogh painting of a woman's head has been revealed with X-ray technology. The painting is thought to have been made in 1884–85, during a period in which he painted several portraits of peasants in the Dutch village of Nuenen. The image was hidden beneath *Patch of Grass*, an unrelated landscape that van Gogh painted a year or two later when living in Paris.

Earlier X-ray studies revealed a faint, blurry shadow of a figure. But Joris Dik of the Delft University of Technology in the Netherlands and his colleagues have extracted a much sharper image using the Deutsches Elektronen-Synchrotron (DESY) in Hamburg, Germany (J. Dik *et al. Anal. Chem.*

doi:10.1021/ac800965g; 2008). The synchrotron's X-ray beam excites secondary X-rays from elements in the sample at characteristic wavelengths. The researchers mapped the distributions of cobalt, arsenic, lead and other metals in the hidden paint layers — all well-known components of pigments that were available at the time. Although the study did not identify all the pigments in the picture, it enabled the researchers to create the partial colour reconstruction shown here.

Van Gogh often re-used old canvases, partly in an effort to save money. Dik's team speculates that he took the portrait with him to Paris, where it would have seemed sombre and unfashionable in comparison to the Impressionists' works, and so he decided to paint a brighter, more commercial floral scene over it. Philip Ball



DELFT UNIV. OF TECHNOLOGY

## Q&A: Pier Oddone

The director of Fermi National Accelerator Laboratory in Batavia, Illinois, talks to **Eric Hand** about the uncertain future of particle colliders in the United States.



FERMILAB

### Did Fermilab welcome the \$337.5 million for science in last month's congressional spending bill after a dismal 2008 budget?

Fermilab received an allocation of essentially \$29.5 million, including \$9.5 million to get Nova [a neutrino project] restarted — it had been zero-ed out in the 2008 omnibus bill. Nova is a project that needs to be done within a reasonable timescale and can't lose much more time. The other \$20 million was used to avoid layoffs and to shore up staff. It eliminated the involuntary stage of our layoffs.

### Delays at the Large Hadron Collider (LHC) at CERN, near Geneva, led to a push to continue running Fermilab's Tevatron. Was that a boon for your scientists?

Our scientists are involved in both colliders, so it's not an issue of it being a boon for us. It has always seemed to me that the prudent thing would be to run the Tevatron until it is clearly overtaken by the LHC, and that

means having colliding beams at energy, with reasonable luminosities.

### The Tevatron is the last US particle collider and will eventually be turned off. What then for Fermilab?

We still have the highest-intensity neutrino beam in the world. We have a project to greatly upgrade that, and it would give us a world-class machine at what we call the 'intensity frontier'. Ultimately, we'd go to Project X and that new beam line would go up to 2.5 megawatts. That's a very rich programme. In addition to doing neutrinos, you can tackle what we call rare decays or rare transitions.

### But many come to Fermilab to work at the energy frontier. What will it mean not to have a collider in the United States?

It would be really depressing if we totally gave up on the energy frontier here. The

energy frontier has moved: for quite a while, it has been in the United States and now it's going to Europe. We have aimed to have a critical mass [of expertise] such that once the detectors are running at the LHC, the experience of coming to Fermilab to do physics will be as rich as it would be to go to CERN.

### US budgets mean that the proposed International Linear Collider (ILC) will probably be hosted outside America.

The method of funding in the United States — year-to-year budgets, with surprises in the middle of a fiscal year — is clearly not a system that would allow you to have a partnership to build an international facility in the United States. It's symptomatic of a problem that I think needs to be fixed; perhaps by treaty, or by putting the money up front, the way we build big aircraft carriers. ■

## Italy launches clinical trial for HIV vaccine

Controversy is surrounding a phase II clinical trial of an HIV vaccine that began in Italy this month.

Led by Barbara Ensoli of the ISS in Rome, Italy's major health research laboratory, the trial is testing a vaccine based on a protein called Tat, which is involved in the replication of HIV.

But critics, including HIV co-discoverer Robert Gallo of the University of Maryland School of Medicine in Baltimore, argue that the scientific basis for Ensoli's vaccine is weak. They note that Tat has proved ineffective in animal models in other labs.

Critics in Italy are also complaining that Ensoli's phase I trial was insubstantial, and that the government has allocated a disproportionate amount of money — €21 million (US\$33 million) over three years — to the work.

Ensoli says that she expects the phase II trial to be looked on favourably by an ISS external committee of experts that will meet this winter.

See also pages 551 and 565

## Nations pool resources to study lunar seismic activity

Representatives of nine national space agencies signed an agreement on 24 July to create an International Lunar Network, which aims to plant a system of six or more seismic stations on the Moon.

NASA scientists joined with space scientists from the United Kingdom, France, Italy, Germany, Japan, South Korea, India and Canada, and agreed to work towards setting up a core set of instruments that could measure 'moonquakes', and the thickness and composition of the Moon's crust, mantle and core.

The programme would begin with a US\$200-million 'anchor nodes' mission in 2014, involving sending a pair of US landers

## Voyage to the bottom of the world's deepest lake

Russian scientists are preparing to dive to the very bottom of Siberia's Lake Baikal — the deepest lake in the world, at 1,637 metres.

The expedition will use the manned submersibles *Mir-1* and *Mir-2* (pictured) and is led by Anatoly Sagalevich, the head of the deep manned submersibles lab at the P.P. Shirshov Institute of Oceanology in Moscow. During around 30 dives, the scientists will take water and sediment samples from Baikal and hunt for hydrothermal vents that spew superheated water into the lake.

Robert Nigmatulin, head of the Shirshov Institute, says that they aim to collect gas hydrate deposits at the bottom of the lake. Gas hydrates have been found in boreholes drilled more than 100 metres below the lake bed and also show up in seismic surveys.

The team plans to explore the lake for about a month and will return to the lake next year.

For a longer version of this story, see <http://tinyurl.com/6jkps8>



Z. DAGBAYEV/ITAR-TASS

to the poles, followed perhaps by a second mission in 2016. The European Space Agency, Russia, Ukraine and Australia are also being included in the discussions.

## GlaxoSmithKline cements interest in stem cells

Drug company GlaxoSmithKline is setting up a US\$25-million, five-year collaboration with the Harvard Stem Cell Institute in Cambridge, Massachusetts. The money will support research as well as a staff exchange programme between the two partners.

Announced on 24 July, the programme is another sign of the pharmaceutical industry's growing interest in stem-cell research, once considered the domain of academics and small biotechnology companies.

The Harvard collaboration is also the latest in a spate of academic collaborations launched by the UK-based drug firm. On 21 July, the company unveiled its 'academic incubator' collaboration with the University of Cambridge, UK, which will give researchers in neuroscience and metabolic disorders a more direct role in the early clinical development of treatments for addiction and obesity.

## Statistics show hint of bias in NIH grant reviews

The system used by the US National Institutes of Health (NIH) to evaluate grant proposals does not adequately compensate for reviewer bias, affecting one in four proposals, a study finds.

Valen Johnson, a biostatistician at the University of Texas M.D. Anderson Cancer Center in Houston, evaluated reviews for

nearly 19,000 grant proposals performed by around 14,000 reviewers in 2005 (V. E. Johnson *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0804538105; 2008).

Each application is typically read by 2–5 reviewers, then discussed when a larger study section of about 30 reviewers meets. In the end, scores from all the study section's members — readers and nonreaders alike — are averaged together.

The system fails to account for individual bias and places undue weight on panel members who have not read the proposals, Johnson argues. He found that the top grants were largely unaffected by reader bias, but that it did affect grants closer to the funding cut-off line. Overall, accounting for reader bias changed about 25% of the funding decisions, meaning that one in four funded proposals would have been replaced by one that had not been funded.

For a longer version of this story, see <http://tinyurl.com/5tlhvj>

## Bail for entomologists held for illegal insect collection

Two Czech entomologists jailed in India on charges of illegally collecting insects were released on bail on 25 July. They are to remain in Darjeeling until their case is heard there on 12 August.

The pair were arrested in late June, for allegedly collecting beetles, butterflies and moths in a national park without proper permits (see *Nature* 454, 14; 2008). Petr Švácha, of the Institute of Entomology at the Biology Centre of the Academy of Sciences of the Czech Republic in České Budějovice, and his colleague Emil Kučera say that to their knowledge they did not even enter the park. If convicted, they face two to seven years in prison.







D. FORBES/FRED HUTCHINSON CANCER RESEARCH CENTER; K. FUNG

# THE NEXT SHOT

Researchers trying to develop an HIV vaccine have endured two decades of setbacks. **Erika Check Hayden** meets a veteran still engaged in the fight — and a rookie willing to join in anyway.

As Larry Corey boarded a plane on the night of 18 September 2007, he was in shock. For four years, he had been in charge of clinical trials of a vaccine against HIV. Although previous vaccines had failed, Corey was optimistic that this one might work because it took a new approach. He thought it was the best hope against HIV. But that afternoon, Corey had met with scientists who had reviewed the early trial results. Not only did the vaccine not work, they told him, it might actually have made some people in the trial, called the STEP Study, more vulnerable to infection.

On the long flight from Chicago home to Seattle that night, Corey cycled through stages of grief, regret and resolve. He had led the search for an HIV vaccine since its inception more than two decades ago. Now, as head of the flagship HIV Vaccine Trials Network run by the US National Institutes of Health (NIH), Corey had to make some difficult decisions. He had to decide what to tell the trial volunteers, and he had to think about the millions of other people who might contract HIV in the future. What could be done to make the vaccine that they so desperately needed?

As Corey pondered these issues, another HIV researcher was mulling over a decision of her own. Colleen Doyle is 32 years younger than Corey and had earned her PhD only three months before. Yet she had just received a job offer from the University of Chicago. The position was a rare and precious opportunity for such a young scientist. But there was a catch: if she took the job, she would have to stop her investigations of how a vaccine could arm the body against HIV.

Both Corey and Doyle (pictured above) had reached turning points, and their decisions could prove crucial to the search for a vaccine. The disappointing failure of the STEP trial marked one of the most difficult moments in HIV vaccine history. It told Corey and other scientists that after decades of work, they still did not understand how to prevent HIV from overwhelming the immune system. And critics were calling for an end to the vaccine hunt, so that some of the money — more than US\$900 million was spent last year (see graph, overleaf) — could

go towards drug treatments.

But where some see an end to the field, Corey and other researchers see a new beginning. They say that basic research will eventually answer the outstanding questions about HIV, if talented researchers like Doyle can be coaxied into the fight. “We need a period of solid, quiet science, even if it takes a decade,”

says virologist and long-time HIV vaccine researcher John Moore from Weill Cornell Medical College in New York.

With this resolution, veteran vaccine researchers are facing a new reality: they will not be the ones to end HIV. They must now pass the baton to a new generation of scientists.

“The real next step is going to come from outside this room,” declared Adel Mahmoud of Princeton University in New Jersey, at an HIV vaccine summit this March in Bethesda, Maryland. But Doyle had watched the first generation of vaccine hunters endure decades of frustration. So the question she faced last September was, did she really want to heed that call?

**“All of a sudden we had this fatal illness, and we were pretty impotent against it.”**

— Larry Corey



Born in 1947, Corey grew up in Detroit and decided at age 10 to become a doctor. As an undergraduate at the University of Michigan in Ann Arbor, Corey started dating Amy Glasser, whose dad had just finished supervising the field trials of the Salk polio vaccine. So when Amy and Larry were married in 1969, Jonas Salk was a guest at their wedding. No one foresaw that Corey would one day emulate Salk's quest to vaccinate humanity against a terrifying viral scourge; at the time, Larry was leaning towards cardiology.

That changed in 1975, when Corey began a fellowship with King Holmes, an expert in sexually transmitted infections at the University of Washington in Seattle, and discovered that he loved research. "That moment of discovery is a wonderful thing," he says, sitting in his office decades later at the Fred Hutchinson Cancer Research Center in Seattle. Now a lanky, grey-haired 61-year-old, Corey seems laid-back, sporting a short-sleeved polo shirt and telling stories about his grandkids. But inside the casual exterior, Corey is fiercely driven, says longtime friend Anthony Fauci, head of the National Institute of Allergy and Infectious Disease (NIAID) in Bethesda, Maryland. "Larry sees his endpoint, he sees his goal, and he goes after it," Fauci says. "He always asks the big-picture questions — he's rarely involved in trivial things."

That was true from the beginning of Corey's career. In 1977, he and Holmes began working with the company Burroughs Wellcome to test a new drug, acyclovir, against herpes. Five years later, their studies led the US Food and Drug Administration to approve the drug, which was directly targeted against a viral enzyme and therefore less toxic than previous antiviral treatments. Barely 35 years old, Corey had already achieved something most scientists never do in a lifetime: the successful development of a drug.

### A fatal disease emerges

The following year, the US Centers for Disease Control and Prevention began issuing bulletins about a disease that would soon be named AIDS. It would later be traced to the human immunodeficiency virus, HIV. As a diagnostic virologist at a regional medical school, Corey began seeing patients with AIDS. Like other doctors, he was troubled by the disease, which did not respond to any known treatment. But Corey at least had experience in developing antiviral drugs, and this led to

his selection, in 1987, as chair of the NIH's AIDS Clinical Trials Group (ACTG).

That year, the first anti-AIDS drug, zidovudine, was approved. But it wasn't a cure, and patients and activists in the United States were growing angry about what they saw as the government's slow response to the disease. As leader of the ACTG, Corey was a frequent target for activist groups, such as the AIDS Coalition to Unleash Power (ACT UP) in New York.

Corey felt the activists' frustration acutely. Each year, he would invite a patient who had contracted AIDS to the first-year medical

school class he co-taught. He would tell students to shake the patient's hand, to touch the patient; that the patient was no different from anyone else, and couldn't infect them with the virus. But every year he had to find a new patient because the previous year's speaker had

died. "In infectious disease, we had the world's best drugs — antibiotics are incredible," Corey recalls. "And all of a sudden we had this fatal illness, and we were pretty impotent against it."

Corey spent five years leading the ACTG. During that time, molecular biologists delivered a nuts-and-bolts understanding of how the virus worked. That enabled them to emulate the acyclovir example and deliver three anti-retroviral drugs designed to fight HIV. Soon, combination therapies would be available for patients with HIV in many wealthy countries. They were sorely needed: AIDS was claiming a shocking and ever-growing human and economic toll. The world needed an HIV vaccine. And Corey wanted to be the man who made it.

**"We need a period of solid, quiet science, even if it takes a decade."**

— John Moore

By 1992, AIDS was the leading cause of death for young adult men in the United States. One of them was Lenny, a distant relative of Doyle's. Both had grown up in a blue-collar community south of Detroit called 'Downriver'. One year, Doyle noticed that Lenny had begun to look sick. But it was only when he was close to death that her parents told her he had AIDS. Even then, Doyle didn't understand why the grown-ups around her seemed so hush-hush. She had seen AIDS on the news and knew that it was a terrible disease that nobody seemed to understand, but she didn't see how it differed from other fatal illnesses, such as cancer.

### Childhood beginnings

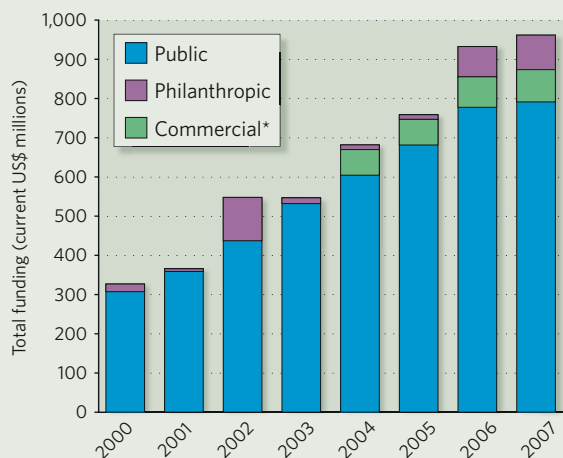
For as long as she can remember, Doyle has wanted to be a scientist. Her favourite childhood toy was a microscope an uncle gave her. She would use it to look at anything and everything, without even really knowing what she was looking for. Then there were the chemistry sets. And the rock polisher. "Science was her own little world," her sister Katie says.

Perched on a stool in her lab, 29-year-old Doyle seems right at home. Clad in jeans and a black hoodie, she pipettes reagents into test tubes sitting in an ice bath while a U2 CD plays in the background. "I've always enjoyed just being in the lab — that's where I feel most comfortable," she says. She joined her first lab in 1997 at the University of Michigan. She wanted to study cancer or HIV, the two big medical mysteries of her youth, but the only lab that would hire her with no experience was a plant genetics lab. She took the job anyway.

The following summer, when Doyle went home, she inexplicably began losing weight. By the time she returned to university that autumn, she had lost 16 kilograms. She was exhausted all the time, her joints were swollen and her hands and wrists would feel frozen. Doctors found that she had an abnormally high sedimentation rate in her blood, often an indicator of inflammation or even cancer. Yet they couldn't find an infection or a tumour. Eventually she withdrew from university and went back home to live with her parents.

Doyle's hometown doctor recognized that her baffling symptoms could be signs of an autoimmune disease. She was referred to a rheumatologist, who quickly diagnosed lupus, an incurable illness caused by the body reacting destructively to its own proteins and DNA. She began taking drugs and avoiding too much sun exposure, as it seemed to set off 'flares'. That same year, her sister was

GLOBAL INVESTMENT IN HIV VACCINE RESEARCH AND DEVELOPMENT



\*Commercial sector investments estimated for selected years

Source: HIV Vaccines and Microbicides Resource Tracking Working Group



**Long shot:** researchers were hopeful that an HIV vaccine could turn T cells against the virus.

P. BRONSTEIN/GETTY IMAGES

diagnosed with ulcerative colitis, an inflammatory disease thought to have autoimmune origins. The sisters then found out that many of their relatives had similar conditions.

By autumn 1998, Doyle felt well enough to return to university and started working in a biochemistry lab. Her passion for science was undiminished. But one thing had changed: she now wanted to study autoimmunity, to understand the illnesses haunting her family.

### Building on a legacy

As the millennium turned, Corey, too, was on a mission. He thought that making vaccines couldn't be too much harder than making drugs and became involved in the first tests of vaccine prototypes. The legacy of Jonas Salk — who had also tried to develop an HIV vaccine — was not far from Corey's mind. Then, in 2000, the NIAID reorganized its vaccine efforts, and Corey won a US\$29-million grant to head the newly formed HIV Vaccine Trials Network, which would plan and conduct clinical trials of vaccine candidates around the globe. The NIAID hoped that its investment in vaccine development — by far the largest in the world — would help make up for industry's indifference.

Salk's legacy wasn't just on Corey's mind — it was about all anyone had to go on. Salk's polio vaccine contained dead polio virus. It trained the immune system to make proteins called

antibodies that recognize the virus and could later be summoned to fight off the live polio virus if it invaded the body. Scientists hoped this 'antibody approach' would work against HIV.

But it didn't. Researchers such as Dennis Burton at the Scripps Research Institute in La Jolla, California, have found that some 'broadly neutralizing' antibodies do block HIV from infecting cells. A handful of people who are infected with the virus but never get sick with AIDS make these antibodies, which recognize proteins in HIV's outer coat even when they mutate to try to evade the immune system. Unfortunately, most people make other, ineffective antibodies that don't recognize the virus when it mutates. In 1998, Vaxgen, a biopharmaceutical firm based in South San Francisco, California, used the antibody approach in the world's first large trial to test whether an HIV vaccine was clinically effective. Vaxgen's vaccine contained pieces of the HIV coat protein called gp120. But that trial and a subsequent one failed miserably. The 7,500 volunteers in the trials did make antibodies to the vaccine, but they were of the weak variety that does not stop the virus<sup>1,2</sup>.

As Corey took the reins of the vaccine network, scientists were changing tactics and turning to another branch of the immune system, the

T cells. They were especially focused on 'killer' T cells, which destroy cells infected by viruses.

In the T-cell game, pharmaceutical company Merck was the leader. The firm had made a vaccine from a common cold virus, the adenovirus, which carried three HIV genes into the host's cells, where they were translated into proteins. Merck hoped that the vaccine would prime T cells to recognize the HIV proteins and destroy cells containing them. The vaccine had looked promising in some tests in monkeys.

The Merck vaccine was the best candidate on the horizon, and Corey was excited about its potential. In 2001, he convinced Merck to form a partnership to test it in a 435-person study that began in 2003. The vaccine generated strong T-cell responses, clearing the way for a second, larger trial in 2004 — the STEP Study. STEP was crucial because it would reveal for the first time whether the vaccine could actually blunt HIV infection. It included 3,000 people in North America, South America, the Caribbean and Australia.

As the trial got under way, Corey was feeling confident. He had maneuvered his network into a prime position at the helm of the world's most promising HIV vaccine trial. It was guaranteed to be a landmark. If Corey's confidence paid off, he would strike a historic blow against a viral foe he had fought for 17 years.

### Keen insight

Doyle's prospects were also brightening. After resolving to study autoimmunity, she had started taking immunology classes and by the time she graduated in 2001, she knew what she wanted to study more precisely than most freshly minted graduates. She was interested in the process that teaches the immune system to

distinguish foreign pathogens from harmless human proteins. Doyle started graduate school that year at the University of Chicago. In 2003, she joined the lab of Martin Weigert, who was doing ground-breaking work on autoantibodies, which attack the body's own tissues.

Doyle's interest in autoimmunity soon steered her back towards HIV. Midway through her graduate work, she noticed a paper from immunologist Barton Haynes at Duke University in Durham, North Carolina. His team had shown that two naturally produced broadly neutralizing antibodies against HIV also reacted against cardiolipin, a molecule found in mitochondria, the cell's powerhouses<sup>3</sup>. The immune system usually has 'tolerance' mechanisms that train antibodies not to attack the body's own tissues. But Haynes thought that these same mechanisms might

**"The field has jumped into vaccine design without really knowing how HIV is working."**

— Colleen Doyle



inadvertently block the production of broadly neutralizing antibodies against HIV that are also autoantibodies — and that might explain why so few people make them.

Haynes's idea made sense to Doyle, who wondered whether her mice made components of autoantibodies so often because some of them could help fight HIV and other infections. "The Haynes paper seemed like a perfect explanation for what we were seeing," she says.

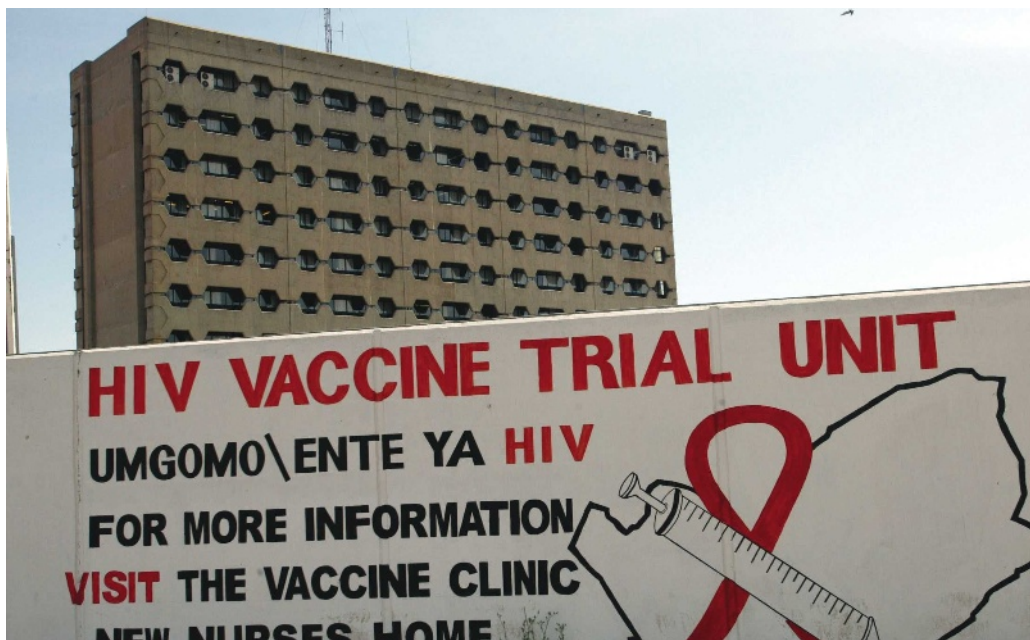
When Doyle began looking for postdoctoral fellowships the next year, she e-mailed David Nemazee at Scripps. Nemazee is an expert in tolerance, and he too was intrigued by Haynes's paper. Nemazee e-mailed Doyle a copy of a grant he was about to submit to the NIH in collaboration with Burton, who worked a floor below Nemazee. The proposal would test Haynes's hypothesis that tolerance prevented the body from making broadly neutralizing antibodies against HIV because they looked similar to autoantibodies. The project seemed perfect for Doyle, offering her a chance to work with two renowned scientists, and to study two problems that had affected her own life. When NIAID approved the \$1.25-million, 5-year grant, she gladly signed on for the project.

### Glowing testimonials

Doyle moved to La Jolla in April 2007, but came back to Chicago in June to celebrate her graduation ceremony. At dinner with her family, Weigert told Doyle's parents how impressed he was with the young scientist. She had been one of his best students: focused and driven, with a gift for experimental design. "He said, 'Do you realize who she is? Do you realize what she's doing, and how important she is?'" Doyle's sister remembers. Once in La Jolla, Nemazee too would become impressed by Doyle because she seemed more confident than most young scientists. "Colleen can think for herself, and ignore you when it's appropriate," he says.

Doyle would need that confidence to see her through the ambitious work that Nemazee and Burton had in mind. The plan was to make mice that carry human genes for broadly neutralizing antibodies against HIV. Doyle would then test whether the mice used tolerance mechanisms to block the assembly of working antibodies. If the mice quashed the broadly neutralizing antibodies, it would suggest that Haynes was right. Haynes, too, was doing experiments to test the theory.

The results should help to explain why the antibody approach to HIV vaccines had failed so far, and inform the design of better vaccine



Thousands of patients have participated in HIV vaccine trials around the world.

candidates. If tolerance mechanisms do block the production of broadly neutralizing antibodies, then researchers might need to turn off tolerance to make a successful vaccine. And even if tolerance doesn't suppress the broadly neutralizing antibodies, the mouse model could still be used to test whether vaccine candidates can stimulate the right antibody responses.

By September, Doyle had almost finished constructing the genetic materials she would use to modify the mice. And that's when Weigert called her out of the blue: just half a year after she had left his lab, a job had come up and Weigert wanted her back at the University of Chicago. He offered her a three-year post studying autoimmunity, one that could become tenured. It would almost certainly end her studies of HIV, yet Doyle found herself leaning towards taking the job. The unexpected

offer just seemed too good to let go.

That same month, Corey too received unexpected news — but his was devastating. "You need to stop the trial," he was told by those on the STEP Study monitoring board. "There's no chance the vaccine is going to work." As scientists analysed the data in the coming weeks, they would confirm that the vaccine had made some people more vulnerable to HIV.

After flying back to Seattle, Corey called Glenda Gray in South Africa and told her to stop STEP's sister trial, Phambili, which had just started in January. The next day, he called Fauci, whose agency helped fund the STEP

trial. "There's no way to torture these data to look good," Corey told his old friend.

As the news reached the rest of the HIV vaccine field, some called it a crisis. Robert Gallo, co-discoverer of HIV, called the STEP results a catastrophe on the same level as the Challenger disaster. Further analysis of the trial results showed that uncircumcised men with strong immune responses to the adenovirus had been made most vulnerable by the vaccine. But researchers didn't know why.

### Early pointers

Critics of the trial said that early tests had predicted Merck's vaccine wouldn't work. They pointed to monkey experiments showing that the vaccine was effective only in animals infected with a weak virus or in those with genetic resistance to the virus. At the annual Conference on Retroviruses and Opportunistic Infections in Boston, Massachusetts in February, Harvard University's Ronald Desrosiers asked an audience of thousands, "Should we be surprised that the Merck product failed to show any efficacy in human trials?" His answer: "Certainly not."

The field needed to understand why the vaccine hadn't worked, and Corey felt the urgency the most. His network had banked on the T-cell approach and planned clinical trials for a roster of new candidates that were similar to Merck's. The most advanced of these was the Partnership for AIDS Vaccine Evaluation (PAVE) trial, which had been scheduled to start in a month's time. PAVE would test whether an injection of HIV genes followed

**"There's no way to torture these data to look good."**

— Larry Corey

J. NGWENYA/IN/WS/REUTERS



by an adenoviral shot could boost T-cell responses in 8,500 volunteers.

After the STEP results were announced, the PAVE trial was put on hold and scaled down. Soon, scientists such as Desrosiers were urging the NIAID to kill it outright: "There is no rational basis for believing that any of the [vaccine] products in the pipeline have any reasonable hope of being effective," Desrosiers said at the February conference. He and 12 other scientists wrote to Fauci, urging him to shift NIH funding towards more basic research. A battle was brewing and millions of dollars were at stake. Should vaccine researchers conduct any more large and expensive clinical trials? Or should they go back to more basic research that might help them design better candidates years down the line?

The arguments came to a head at a March summit organized by the NIAID. Like Desrosiers, many of the 200 or so vaccine researchers at the meeting felt that only basic research could answer the questions that had stymied the search for an HIV vaccine. If they were to pursue the T-cell approach, they needed to know what type of T-cell response could stop the infection from establishing a fatal foothold in the body. If the T-cell approach was doomed, they had to resurrect the antibody approach, and find out if there was any way to coax the body to make neutralizing antibodies against the virus.

### Money matters

Their arguments won the day. Fauci agreed at the meeting that the NIAID would shift more money to basic research, which at the time claimed just under half of the agency's \$476-million vaccine budget. After the summit, the NIAID made plans to move \$50 million from product development to basic research over the next few years, starting with \$17 million in 2008. In May, it started out by giving \$4 million in redirected dollars to a \$15.6-million programme to fund research on B cells, which make antibodies. Not that clinical development was written off: dozens of phase I and II clinical trials are planned or under way, some funded by NIAID and other government sources, and some backed by private groups.

Then, on 17 July, Fauci announced that the PAVE trial would be cancelled. The \$140 million that would have funded it would be scaled back for a much smaller study. The decision would cut the budget of Corey's network but would free up between \$5 million and \$10 million for basic research. "The differences

between STEP and PAVE were not so great as to warrant doing this trial," Fauci said.

The decision will shift the vaccine network's emphasis towards basic research — a move that Corey agrees is necessary. He hopes that this will lead to more promising clinical trials in the future. But when those studies happen, Corey himself might not be there to conduct them. Being the public face of the vaccine effort, he says, can be "more of a burden than a joy".

Making a vaccine against HIV has been so



Anthony Fauci's agency devoted more money to basic research after the failed vaccine trial.

much harder than Corey ever imagined, and it may not even happen within his professional career. But Corey knows that he has been fortunate to be a scientific researcher. "You're given money by the US taxpayers to follow your own ideas," Corey says. "That's a privilege." Even if no vaccine is in hand, Corey hopes that the money has delivered some valuable lessons about HIV. But sometimes he wonders whether he should

return to the basic molecular virology work that was his first love. "I still hope I am worthy of that," he says.

And while senior researchers at the crisis summit were crying out for basic research and more young scientists, Doyle was quietly getting on with her work. She had made her decision about the job

offer in Chicago: she would turn it down. She had become so interested in her new project on HIV that she realized she didn't want to quit the field, even if it meant giving up an opportunity: "I decided I really wanted to work on HIV vaccine design, and this is the best place right now for me to do that," she says.

Doyle had heard about the failed STEP

trial and its dramatic aftermath. But it hadn't deterred her. "The field has gotten ahead of itself, and jumped into vaccine design without really knowing how HIV is working," she says. "A lot of work needs to go back to figuring out what makes this virus different."

Doyle hopes that her own work will aid that effort. And so does the NIAID: on 20 May, the agency announced that Nemazee and Burton will receive one of the back-to-basics grants on B-cell biology awarded in the wake of the vaccine summit. The \$2-million grant will support studies on autoimmunity and neutralizing antibodies, including Doyle's project. At the moment, Doyle is still trying to breed mouse pups carrying genes for the neutralizing antibodies.

### Progressive steps

There are already signs that she might be on the right track, however. Doyle has also made mouse B cells that carry the genes for one of the HIV-neutralizing antibodies. To her surprise, she found that the cells react with DNA. That's intriguing, because antibodies that react with an animal's own DNA are, by definition, autoantibodies. Nemazee is sceptical though, because earlier data told a conflicting story, and Doyle's Petri-dish studies might not predict what actually happens in a human being. He says that Doyle's mice will be a more convincing test of the autoantibody hypothesis.

Doyle's studies may or may not hold the key to an HIV vaccine; basic research never guarantees an answer. But Doyle is exactly the kind of person who should be asking the questions. Young, sharp and fascinated by biology, she is motivated by the same thrill of discovery that brought Corey into science. She, too, knows that she is privileged, even if that privilege involves countless afternoons, like the one she spent earlier this month, sitting alone in a darkened basement, using a cell sorter to analyse tubes full of engineered mouse B cells.

Doyle watched screen after screen of data flash before her eyes on a computer. The data wouldn't yield a magic bullet that day, or possibly ever. But she and many others in the field think that it is work like this in labs around the world that holds the only hope for an HIV vaccine. In the long quest for that vaccine, Corey and his generation of researchers have offered a baton to Doyle and hers — and Doyle, for one, has taken it, and is off and running. ■

**Erika Check Hayden is a senior reporter for Nature based in San Francisco.**

1. Flynn, N. M. *et al.* *J. Infect. Dis.* **191**, 654-665 (2005).
2. Pitisuttithum, P. *et al.* *J. Infect. Dis.* **194**, 1661-1671 (2006).
3. Haynes, B. F. *et al.* *Science* **308**, 1906-1908 (2005).

See Editorial, page 551.

**"Should we be surprised that the Merck product failed to show any efficacy in human trials? Certainly not."**

— Ronald Desrosiers



# UPGRADING THE GRID

Electricity grids must cope with rising demand and complexity in a changing world.

**Emma Marris** explores the intricacies involved in controlling the power supply.

**O**n a hilltop in the small city of Vancouver, Washington, just across the Columbia river from Portland, Oregon, sits a concrete building owned by the the Bonneville Power Administration (BPA), the federal agency that runs the electricity grid in the Pacific Northwest. In that building's basement one can find Albert Orona enthroned behind a bank of monitors, facing a wall-sized map of the region's power grid.

Orona is a dispatcher, one of the unseen men and women who keep power moving smoothly through the grid. It is not a job that would even occur to most users. Electricity is taken for granted: just flip a switch and the light goes on. But to Orona, contending with a network of almost unfathomable complexity, electricity is anything but a given.

From his basement command post he oversees more than 24,000 kilometres of high-voltage wires channelling power from one nuclear power plant, several coal plants, 31 dams scattered throughout Washington, Oregon, Idaho and Montana, and an ever-increasing number of wind farms. And that is just the 'bulk transmission grid'; the local distribution from substation to customer is handled separately by the utility companies, the people you send your money to each month.

In Orona's domain, something always needs adjusting. When a generator goes down for scheduled maintenance, or fails, he and his team will call for other generators to increase their output. If a disturbance in the grid is imminent, Orona will 'island' parts of the system to isolate them from power fluctuations

that could otherwise cause a blackout. And if a blackout threatens to spread — well, it is hard to imagine the calmly competent Orona discombobulated. "If I was a kid," he says, "I would love to do this for free. It is a fun job." But in a crisis, he admits, "there is a certain adrenaline flow".

That adrenaline is flowing through the whole electric-power industry these days. Demand for electricity is soaring worldwide. And yet, at least in developed countries such as the United States, a combination of high costs, environmental concerns, and uncertainty over post-Kyoto carbon regulation is making it harder and harder to build new power plants or run new transmission lines. So the grid is increasingly run on the ragged edge of failure, flirting dangerously with 'unstudied states' — situations in which, as Orona puts it, "if I lose another element, that might lead to cascades that would take out that part of the world".

**"If I lose another element, that might lead to cascades that would take out that part of the world."**

— Albert Orona

The challenges that Orona faces at Bonneville are all too typical of these global issues. Fortunately for him and for others in his position, research is now being conducted on ways to hold those unstudied states at bay. New tools include sensors and other 'smart-grid' technologies that will make the grid better able to manage itself. They include energy-smart appliances, local wind- and solar-power generators, as well as 'demand-side' technologies that will help consumers control how much power they draw from the grid.

And, of course, they include measures to get both approaches — smart grids and demand management — working together. After all, the

more that sources of power generation diversify, and the more information that clever appliances send back up the wires, the smarter the grid will have to be to cope with the complexity.

The result could be something completely new. If today's grid is Hollywood in the 1930s — with a few big studios piping content for viewers to watch passively in a theatre — tomorrow's grid will be YouTube, with thousands of smaller players and the line between consumer and supplier considerably blurred.

## Mapping the system

One of Orona's colleagues is on a ladder updating the aqua-tiled wall display with paper status tags. The board shows the grid like a subway map. A lit bulb indicates that a plant or substation is offline. And black dots indicate power plants that can 'black start', or begin generating power from scratch without any power flowing in. Meanwhile, Orona is sipping his soft drink and flipping through different screens on his seven monitors. Alarms are going off at the rate of about two a minute. One sounds like a van backing up, another like the wrong answer on a game show. Some Orona seems to ignore, others cause his head to flick momentarily to one monitor or another.

There are three fundamental facts that ensure Orona's job is never dull. First, electricity moves at nearly light speed, so that transmission is essentially instantaneous; the electric field illuminating your light bulb this second was born in a dam or some other generator this same second. Second, electricity cannot be stored, except on a very limited scale. Taken together, these two facts mean that Orona and his counterparts elsewhere have to keep generation almost exactly balanced with



demand on a second-to-second basis. When a baseball stadium flips on the lights in Seattle, the Grand Coulee Dam in eastern Washington has to route a bit more water to its turbines.

Except that it isn't that simple — which is the third fundamental fact: the grid is extraordinarily complex. The power flowing through the stadium lights cannot actually be traced back to a single generator; it comes from the common flow of energy fed by all the generators. (Think of dipping a bucket into a reservoir fed by many rivers: it pulls up water from all of them.) So the grid has to be managed as an integrated whole. Yet the grid is also a hotchpotch: in much of the world it has been 'integrated' over the decades by patching together small, local grids as the opportunities arose. And the energy pulsing through that hotchpotch can be downright willful. If a transmission line goes dead, the electricity will spontaneously reroute itself along any other path it can find. So if there aren't a lot of redundancies in the system — as often happens these days — and if the extra power moves to other lines that are already near capacity, those lines might also overload and shut down. This can lead to still more shutdowns, in an ever-increasing chain reaction that becomes a region-wide blackout.

Or maybe not. Anticipating every conceivable sequence of failures is all but impossible in a system this complex. That's why Orona talks about unstudied states. Even the grid's normal operation is difficult to predict. All the current computer simulations are lousy, says Bonneville modeller Scott Simons, despite ongoing efforts to improve them. "We are getting pretty good for one hour from now, not so good for two hours," he says. "Three hours is pretty bad."

In short, humanity has come close to building a machine that is so intricate that it can't be comprehended. But Orona and his fellow power dispatchers have to manage that machine

anyway; modern society depends on it.

Given the increasing complexity of the grid, the most immediate priority is to give the dispatchers an upgrade in the 'situational awareness' department. Situational awareness, a favourite phrase of grid managers, means 'knowing what is happening'. Bonneville is already working to modernize all those displays and alarms. For example, since May, when the visit described in this article took place, the aqua-tiled wall display has been replaced with an electronically updatable digital version. Efforts are also under way to improve the user interface so that Orona can get the same data about the grid on fewer monitors.

### Web of information

Out in the field, next-generation sensors will soon be feeding Orona better data. New 'synchrophasor' units, for example, use the precision

of the clocks on Global Positioning System satellites to compare two frequency measurements taken in different parts of the grid at the exact same time. This allows them to hunt for sudden

changes that hint that the grid is under a lot of stress, or that something major has gone out and the rest of the grid is shifting accordingly. Compared with the grid's older mix of analogue and digital sensors, says Carl Imhoff, an engineer at the Pacific Northwest National Laboratory in Richland, Washington, "phasor networks sample the grid 30 times a second instead of once every 6 seconds".

Meanwhile, improvements in real-time simulations of the grid are helping dispatchers to make better use of that information. Simons is working on one such computer model that will provide operators with the most economical mix of generation to meet demand at any given moment.

The cheapest power on the grid isn't always the best deal, because power

**"The grid is always changing, and it is almost self-correcting."**

— Jon Ludwigson

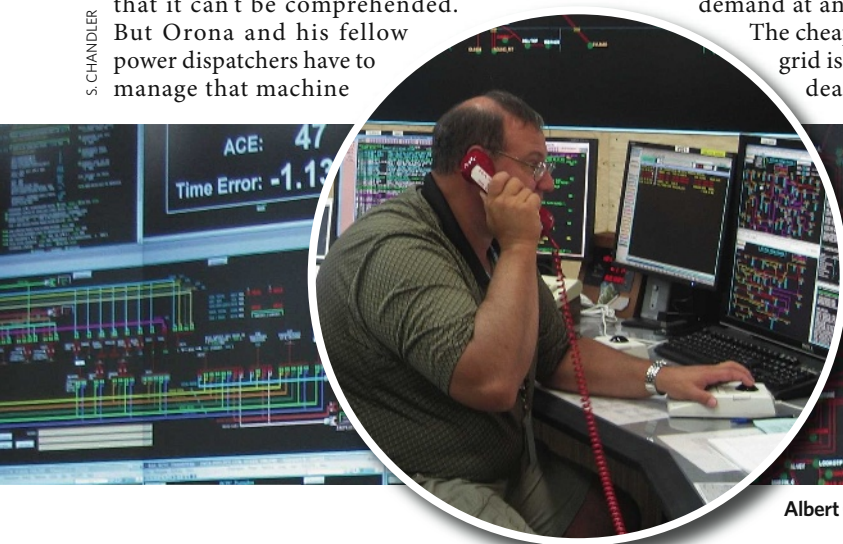


KAJ R. SVENSSON/SPL

Wind power is now an important electricity source.

leaks out of the lines at a rate that depends on distance and wire type. According to Simons, the model even takes into account whether certain efficiency gains are too puny to be worth a dispatcher's time to make the requisite phone call. "If we called Grand Coulee Dam and said 'tell you what, move your generation up two megawatts' [about 0.03% of the dam's total capacity of 6.8 gigawatts] the laughter would deafen you," says Simons.

Still, despite the computer programs, and the efforts to reduce the cognitive load, Orona's chief tool in managing the grid continues to be a red telephone of the direct-line-to-the-Kremlin variety. When there is a problem, or even the hint of a potential problem, Orona gets on the red phone and, for example, asks another utility to make an adjustment. In the not-too-distant future, however, his part of the grid may very well be able to talk to that part of the grid without human aid. Indeed, there has been a global resurgence of efforts to move most or all of the management of the grid into the grid itself. Under its €2.3-billion (US\$3.6-billion) Seventh Framework Programme, for example,



Albert Orona on the red phone in the Bonneville Power Administration control room.

the European Commission is planning to fund a number of projects in smart energy networks.

And the recent US Energy bill, signed into law on 19 December 2007, calls for a 'Smart Grid Task Force' to run up to five demonstration projects focused on smart grids and energy reliability, and calls on the National Institute of Standards and Technology to develop a set of standards that will allow all such smart grids to be interoperable.

### The self-managing grid

In some ways the grid is already intelligent. No human needs to trigger a circuit breaker on an overheated transmission line; that happens automatically. "The grid is always changing, and the way it is designed it is almost self-correcting," says Jon Ludwigson, a grid expert at the federal government's General Accountability Office in Washington DC. "If a power line goes out, the grid still flows."

To add to this, the controllers are aided by programmed remedial action schemes, also called special protection schemes. These are rules that will automatically trigger an action when a particular threshold is reached. Imagine, for example, that a significant transmission line is lost on a congested route over a mountain pass. To avoid a cascading outage, the reprogrammed routine might automatically cut back on power generation and, to keep things in balance, reduce the load on the grid by shutting off a few power lines downstream of a substation. A few blocks of one town might go black, but the rest of the towns won't even know what a close call they had.

"We stretch the power lines to the limits, and the remedial action scheme allows us to run it at the limits," says Orona.

But in a truly smart grid, the analogous programs would run in processors attached to the various key components of the grid, and would be fed by a rich stream of real-time data flowing in from sensors all over the grid. In principle, says Massoud Amin, an electrical engineer at the University of Minnesota in Minneapolis, such a smart grid could be not only self-managing, but "self-healing" (see graphic). In his vision, he says, the processors in the future grid will be able to "localize and anticipate the consequences of disturbances, whether they are natural disturbances, such as lightning or hurricanes, or intentional disturbances". By the time the wind-snapped or sabotaged line hits the ground, the whole grid will have shimmered into a new configuration to stymie disastrous cascades. "Electricity travels almost at the speed of light," says Amin, "so we have a few milliseconds to take this action before it becomes widespread." He estimates that such a grid in the United States would take ten years to roll out and cost between \$10 billion and \$30 billion a year to install, shouldered by a public-private partnership. That's no budget operation but, according to Amin, it would cost just a seventh or an eighth of the current annual cost to society of power interruptions.

Many grids outside the United States are further ahead in the intelligence-raising process,

with particularly advanced projects in Italy, Sweden, and in the state of Victoria in Australia. Enel, Italy's largest utility company, replaced 30 million old power meters with smart, microprocessor-equipped meters that have made possible new pricing structures that encourage customers to shift their power usage to off-peak hours.

However, no matter how smart the grid becomes, it will eventually be overwhelmed if demand keeps rising. So industry planners are also moving to embrace micro-generation and to reduce demand. "In the old days we kept the whole thing balanced with one wrench: the supply side," says Imhoff. Being able to tweak the system from the demand side will be a huge improvement; it will also make the

**"Electricity travels almost at the speed of light, so we have a few milliseconds to take action."**

— Massoud Amin

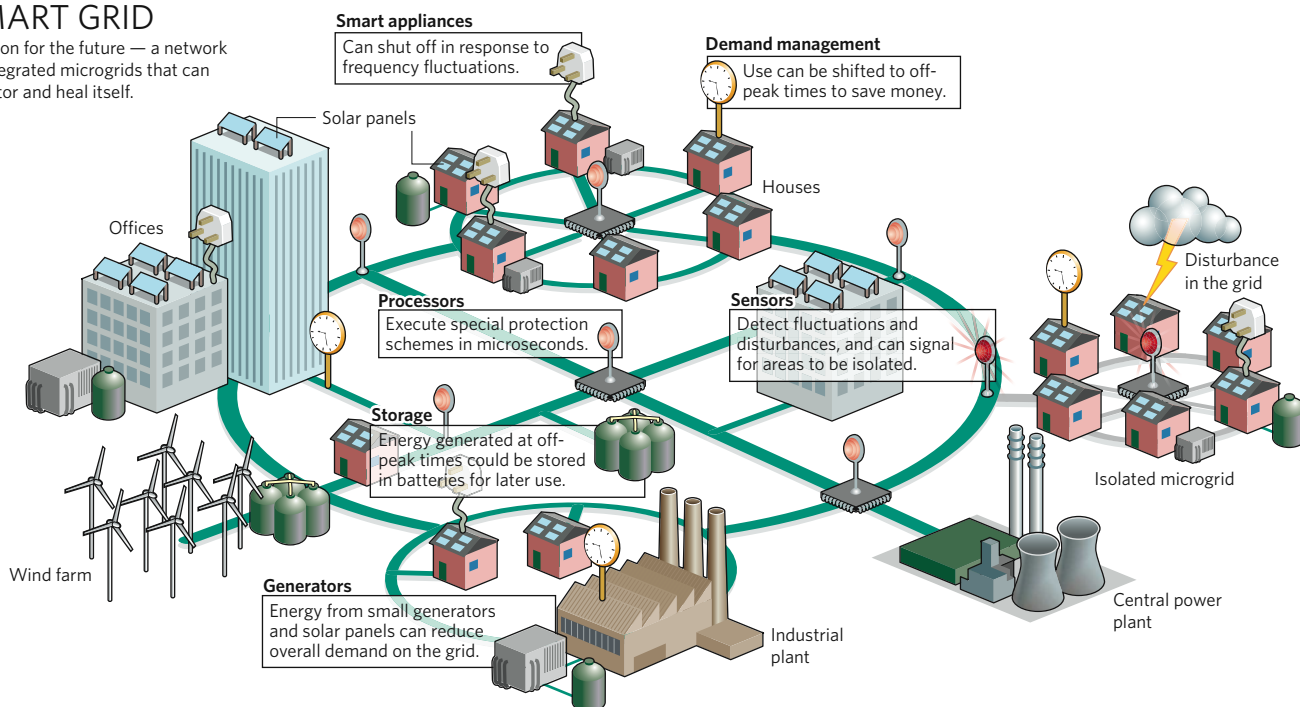
grid that much more complex.

Currently, most consumers suck up electricity in a predictable and mostly unconscious pattern. In the morning, people turn on lights and radios, use kettles and toasters, adjust thermostats and check e-mails. Spike number one. Then people head to work or school, and the electricity dispatchers can have a cup of coffee. After the mid-day lull, the pattern reverses. People drive home, turn on the heat or air conditioning, cook dinner, do the laundry, surf the Internet and watch television. Spike number two.

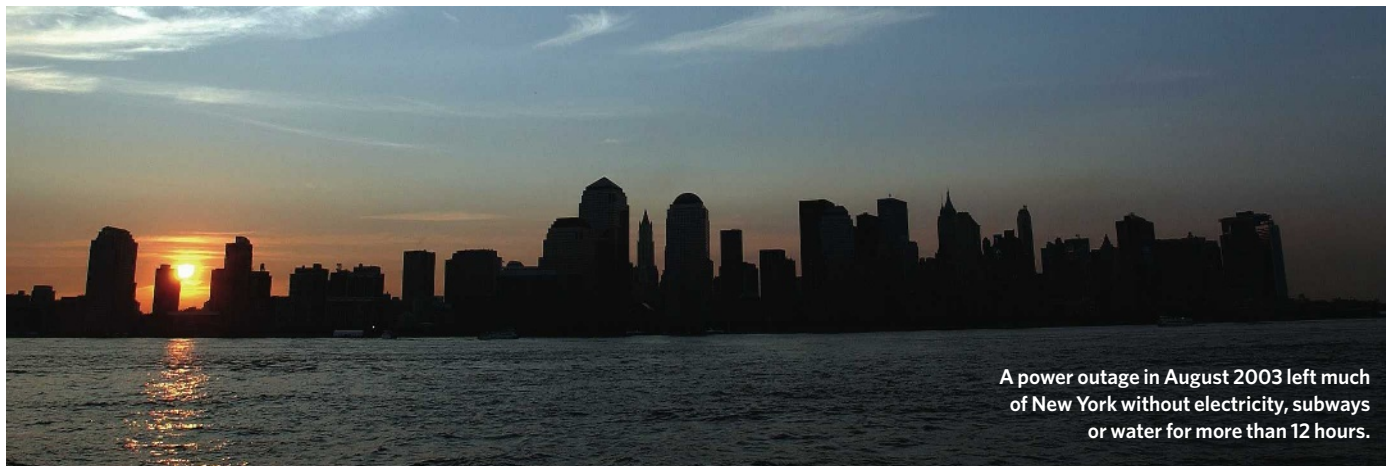
But if electricity were to cost a lot more during peak hours, and if people's appliances could tell them so, perhaps that unconscious pattern would change. Then again, maybe it wouldn't be enough. Until it is tried, all anyone can say

## SMART GRID

A vision for the future — a network of integrated microgrids that can monitor and heal itself.







REUTERS/CORBIS

A power outage in August 2003 left much of New York without electricity, subways or water for more than 12 hours.

for sure is that the usage peaks would be less predictable. Price fluctuations would ripple because of people's behaviour and then price signals would race backwards towards buyers and sellers in a highly nonlinear set of feedback loops. Small home generators might simply cut off the top of peaks of demand — or they might turn the peaks into troughs, by sending extra electricity into the grid. Suddenly, the game would have a lot more players.

### Supply and demand

The first steps towards this future may look familiar and reassuring to consumers. At the Pacific Northwest National Laboratory (PNNL) in Richland, Washington, a decidedly normal-looking demonstration kitchen sits in the corner of a lab. But if the stolid white refrigerator detects a disturbance in the grid — as evidenced by frequency fluctuation — it will shut off for a few seconds to shed load. Crucially, however, it will not shut off power to the circuit for the little light bulb inside the fridge; the customer will never notice that the appliance is off. Likewise, the clothes in the dryer will keep tumbling even though the heat has been momentarily shut off. Recently, the PNNL tested 150 such dryers in homes in the area. Consumers didn't notice a thing and said they'd be happy to buy them. The researchers think that if the whole country has such appliances, 20% of national demand could be put on hold at any moment.

But that is just the beginning. The PNNL is also working on dynamic pricing and distributed generation. Many commercial customers in California already have the option of saving money by running their machinery at night, but PNNL is examining how real-time price fluctuations could be extended to every end user. In a pilot project on the Olympic Peninsula of Washington state that ran from March 2006 to March 2007, residential, commercial and municipal customers in this mountainous

region could track prices on their computers, with updates every five minutes. Residential users could set their thermostats, dryers and heaters to respond to certain price points. Small generators — for example, backup diesel generators for a municipal water-pumping plant — were programmed to kick in when grid power got too expensive.

In the end, the experiment smoothed out grid congestion and customers saved, on average, 10% on their electricity bills. One enthusiastic participant, Jerry Brous, wrote a memo for the project's final report about paying attention for the first time to how he and his wife Pat use electricity, and discovering just how cold or hot they could tolerate their house and water. "It is also great fun," Brous wrote, "to sit at a picnic table at an RV park and jump online through a Wi-Fi connection and tell the water heater and heat pump in our house to wake up and get to work, we are coming home early."

Similar projects around the world have generally been received favourably by participants. If such technology is deployed widely, then in 20 years the average residential consumer might well be much more strategic and conscious about their use of energy. That prospect becomes even more likely if small household generation with solar panels or other renewables becomes commonplace; coupled with smart-grid technology in the distribution network as a whole, this would enable neighbourhoods inside a city to 'island' themselves from the wider grid when there is a disturbance.

Another piece of the puzzle is provided by energy storage. In Japan, to take just one example, the Tokyo Electrical Power Company has co-sponsored the development of large commercial sodium-sulphur batteries that can hold electricity made in off-peak hours for deployment during peak demand times.

Denmark, which produces more wind power per inhabitant than any other country, is likewise looking into storing off-peak energy in vanadium-based fuel cells.

Of course, none of these developments is going to put Orona and his team out of work anytime soon; some cognitive leaps are still beyond computers. "Sometimes you can have a problem that occurs because you have bad data," says Orona. "Only the human can say, 'Hey — those are bad data.'"

When Bonneville engineer Lawrence Carter was asked how long it would take for a major blackout to develop if all the dispatchers at Dittmer and other operations centres suddenly disappeared, he took a while to answer. "Eventually, the system would fault," he said at last. "It can't run itself." A high-speed computer, Carter says, would take three years to compute every single fault path in the system. But a human is smarter. "People learn about their paths," he says. "You may not have to do those 30 million simulations; you might be able to just say, 'This is going to go next.'"

Still, there is no doubt that the future of the grid will involve an ever-expanding cornucopia of new technologies: smart appliances, dynamic pricing, micro-generation, energy storage, built-in protective responses — on and on. "No one by themselves is going to help us address all the challenges that we have," says Gil Bindewald, acting deputy assistant secretary for research and development at the Department of Energy's Office of Electricity Delivery and Energy Reliability. "You might have underground cables in one place, overheads in another, a plug in hybrid there, a microgrid there. I think the entire grid and how we interact with it will change. It is just not the socket in the wall."

Emma Marris writes for *Nature* from Columbia, Missouri.  
See Editorial, page 551.

**"The entire grid and how we interact with it will change."**

— Gil Bindewald

# CORRESPONDENCE

These letters respond to the Commentary 'Repairing research integrity' by Sandra L. Titus, James A. Wells and Lawrence J. Rhoades (*Nature* **453**, 980–982; 2008).

## Integrity: Croatia's standards unusual in much of Europe

SIR — Sandra Titus and her colleagues argue that a failure to foster a culture of research integrity is the common denominator in scientific misconduct, which in the United States is much more prevalent than might be expected. But what would the results have been if the survey had been done in countries without the codes of good practice and procedures for handling allegations of misconduct laid down by the US Office of Research Integrity (ORI)?

Take Europe, where — apart from in Scandinavia, Germany, the United Kingdom and, to some degree, France — little or no regulation exists to control scientific misconduct. Individual cases of fraud can therefore be more easily hidden and may be far more common than in countries with established standards.

Horace Judson suggests that fraud is intrinsic in cultures "characterized by secrecy, privilege and lack of accountability" (*The Great Betrayal: Fraud in Science* Harcourt, 2004). These features are evident in some Spanish and Italian research organizations, for example, where cronyism is rife and transparency is obscured. Russia allegedly maintains a tolerant attitude towards violations of medical ethics, and Marek Wroński claims, perhaps too strongly, that an "old boys' network" protects scientists from accusation or prosecution in Poland (M. Wroński *Przegl. Lek.* **55**, 629–633; 1998).

In any case, the real extent of misconduct in Europe is largely unknown and inadequately investigated.

Countries from southern and eastern Europe, say, could well

emulate the standards already in place elsewhere. In Scandinavia, for example, training in good research practice is mandatory for researchers, as it is in the Biomedical Research Park of Barcelona in Spain. Croatia deserves special mention: its leading medical journal has been cooperating with the ORI since 1999 and has initiated a dialogue on research integrity with the Croatian Science Ministry (M. Petroveci and M. D. Scheetz *Croat. Med. J.* **42**, 7–13; 2001). The journal has taken the unusual step of creating a 'research integrity' editor, and the country has been actively teaching responsible research conduct since 1996.

Punishment makes no sense without prevention, and prevention is necessary because, as noted by Titus and colleagues, self-regulation is unlikely to be effective.

**Xavier Bosch** Department of Internal Medicine, Hospital Clinic, Villarroel 170, 08036 Barcelona, Spain  
e-mail: xavbosch@clinic.ub.es

## Integrity: juniors see leaders gain from calculated dishonesty

SIR — The survey by Sandra Titus and her colleagues of extramural scientists funded by the US National Institutes of Health indicates that many cases of scientific misconduct may not be reported to their institutions and that even fewer come to the attention of the government funding agency. They point out that institutional leaders have a strong financial incentive to silence the would-be whistle-blower, because a confirmed case of research misconduct can hurt the institution's reputation and impede the flow of its funding.

But in calling for more tutoring and stricter supervision of junior

scientists, Titus and colleagues drastically oversimplify the nature of the problem. They assume that most scientific leaders behave with integrity in their own work, but somehow fail to pass this trait to those whose research they supervise.

On the basis of our own discussions with biomedical scientists at the predoctoral, postdoctoral and faculty level, we hold a different view. The academic and financial rewards of calculated, cautious dishonesty on the part of some scientific leaders are, we believe, all too apparent to the junior scientists they supervise. No amount of tutoring, stricter supervision or courses in research ethics will fix this problem.

We, the writers of this Correspondence, are the authors of a report written 21 years ago on misconduct in biomedical research (*Nature* **325**, 207–214; 1987).

**Ned Feder** Project On Government Oversight, 666 11th Street NW, Washington DC 20001, USA  
e-mail: nfeder@pogo.org  
**Walter W. Stewart** Lone Rider Trail, Austin, Texas 78738, USA

## Integrity: misconduct by a few damages credibility for many

SIR — It is disturbing to read about the prevalence of misconduct described by Sandra Titus and colleagues. But, as scientists, we are trained to deal with reality, not to avoid or, worse, misrepresent it. The description of a scientist as

honest should be a tautology.

An institution can keep its good name only by reacting vigorously against allegations of misconduct. It should publicly denounce practices such as data falsification, plagiarism and duplication of research results in different publications. Official statements should be issued, warning that misconduct allegations will be subject to formal investigation and the results made public.

Unfortunately, here in Brazil this is not common practice. If misconduct allegations are ever filed, official statements are usually vague and investigations can take several years. Whistle-blowers are typically frowned upon by their colleagues and officials at their institutions.

Even though we agree that regulatory offices cannot catch all misconduct events, we suggest that consideration be given to the creation of international 'offices of research integrity' to pursue universal standards of ethical behaviour. After all, the unethical behaviour of a few scientists can damage the credibility of many.

**Paulo A. Nussenzeig, Renata Zukanovich Funchal** Instituto de Física, Universidade de São Paulo, PO Box 66318, São Paulo, SP 05315-970, Brazil  
e-mail: nussen@if.usp.br



J. TAYLOR



## Integrity: how to measure breaches effectively

SIR — When suspected scientific misconduct occurs in a research department, it is likely that more than one person knows about it. In their Commentary, Sandra Titus and colleagues avoid the multiple-reporting problem in estimating the incidence of misconduct by surveying one person per academic department about suspected misconduct within that department. However, I question their extrapolation of these survey results, which they claim projects an alarming picture of under-reporting.

The authors derive a rate of 0.03 cases of suspected misconduct per department per year, but settle on a more conservative figure of 0.015. They then apply this rate to the total population of 155,000 researchers funded by the US National Institutes of Health (NIH), arriving at an extrapolated estimate of a minimum of 2,325 cases of suspected misconduct per year.

It is not appropriate to extrapolate from a sample of departments to a universe of individuals. Applying the 0.03 rate to a rough estimate of 10,000 departments with NIH funding, the authors could claim an extrapolated estimate of only 300 cases of suspected misconduct per year.

Titus and colleagues cite our earlier study (J. Swazey, M. Anderson and K. Lewis *Am. Sci.* **81**, 542–553; 1993) as methodologically weak in its estimate of misconduct incidence, because we allowed multiple reports within departments. The difference is that we neither aimed nor claimed to measure incidence, but rather to measure scientists' exposure to suspected misconduct.

The authors' extrapolation seems, like ours, to estimate exposure and not incidence.

**Judith P. Swazey** PO Box 243,  
Bar Harbor, Maine 04609, USA  
e-mail: jswazey@verizon.net

*Titus et al. reply:*

As Swazey points out, the earlier study (J. Swazey, M. Anderson and K. Lewis *Am. Sci.* **81**, 542–553; 1993) measured exposure to misconduct, rather than its incidence, when multiple respondents in the same department were reporting on the same case. The only circumstance in which exposure points to incidence is when there is only one observer per unit of observation. We therefore designed our study so that we sampled only one scientist per department.

Swazey's statement, however, implied that our results should be extrapolated only to departments. Her comment assumes that each observer is reporting all incidents in the department, rather than just those that he or she observed. This is unlikely even in a moderately sized department, let alone in a very large one. In a separate analysis (Gallup Organization *Final Report: Observing and Reporting Suspected Misconduct in Biomedical Research* Washington DC; 2006, table 8, see ORI.hhs.gov/gallup08), we showed that the incidence of reporting was not affected by department size, and so justified an extrapolation based on the number of research personnel supported by the National Institutes of Health.

**Sandra L. Titus** Office of Research Integrity, 1101 Wootton Parkway, Rockville, Maryland 20852, USA  
e-mail: sandra.titus@hhs.gov  
**James A. Wells** Office of Research Policy, University of Wisconsin-Madison, 205 Bascom Hall, 500 Lincoln Drive, Madison, Wisconsin 53706, USA  
**Lawrence J. Rhoades** Office of Research Integrity, Maryland 20852, USA

## Digital identifiers could keep up with authors' moves

SIR — *Litera scripta manet:*

'written words will endure'.

But not, it seems, in the case of the e-mail addresses of corresponding authors in the scientific literature.

To investigate the survival rate of author e-mail addresses, I sent an e-mail to the first one hundred corresponding authors of peer-reviewed papers whose addresses were returned in a Google Scholar search for 2007 and 2003. Roughly one out of five messages was undeliverable in 2008 (from 2007: 17%; 2003: 25%), indicating that those e-mail addresses were no longer valid.

E-mail addresses of scientists, particularly those without tenure, are volatile. Researchers leave behind a trail of obsolete e-mail addresses, phone numbers and fax numbers in the printed literature.

Unique digital author identifiers, as proposed in Correspondence (*Nature* **453**, 979; 2008), could be linked to up-to-date e-mail addresses and other contact information. This would increase the traceability of authors, facilitate scientific networking, and even speed up the peer-review process.

**Raf Aerts** Division of Forest, Nature and Landscape, Katholieke Universiteit Leuven, Celestijnenlaan 200E bus 2411, BE-3001 Leuven, Belgium  
e-mail: raf.aerts@ees.kuleuven.be

## Schools in a third of Spain teach only in minority languages

SIR — Your Editorial 'Comédie française' (*Nature* **453**, 1144; 2008) argues that opposition by the members of the *Académie française* to including regional languages in the French constitution is disingenuous. But maybe these French

academics have looked south and seen what has happened in Spain, where "regional and minority languages, like endangered species", are considered to "merit protection" by several of the regional governments.

Today, it is impossible to obtain public or publicly funded education in Spanish, the common language, in the schools of about one third of the country, including Catalonia, Mallorca and Valencia. For example, teaching is conducted in Catalan or one of its variants in northeastern Spain, and in Gallego in Galicia in the northwest.

In the Basque country, despite the obscurity of the language, education programmes will be available only in Basque from 2009 and programmes taught partially in Spanish will be dropped.

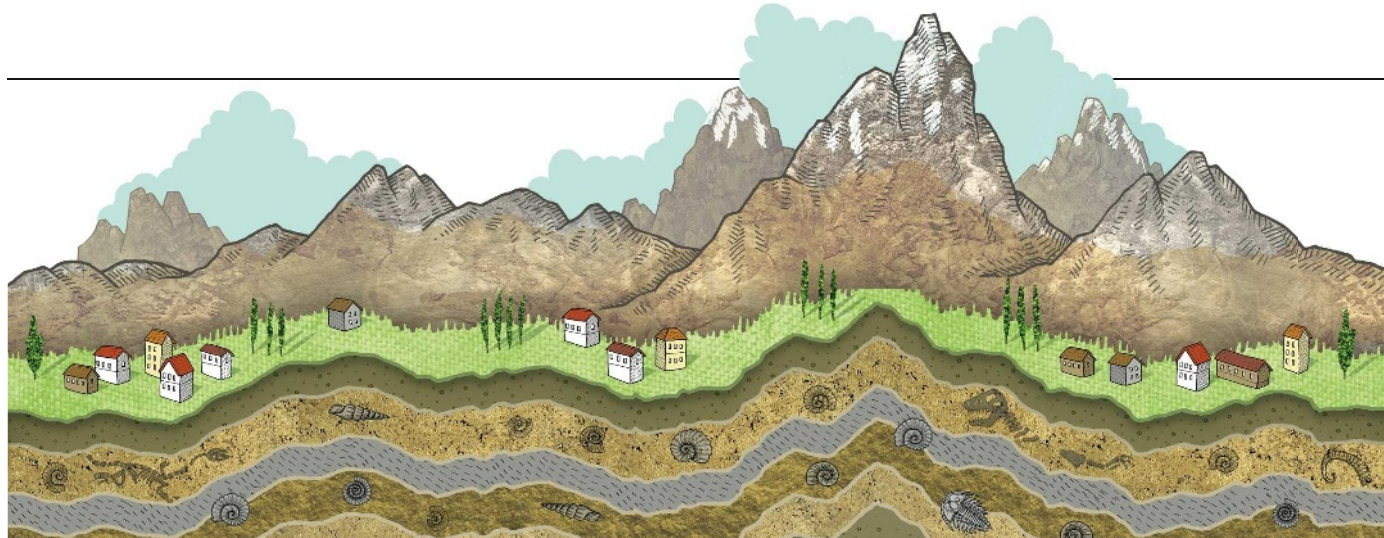
This is an absurd situation, where in some places it is easier for Spanish children to study in English (for example, in the British Council schools) than in Spanish, the language that the Spanish constitution has set as the common official language.

It has stimulated prominent — and by no means all conservative — intellectuals, headed by the novelist Mario Vargas Llosa, to sign a manifesto calling to defend the rights of Spanish-speaking people in their own country (see <http://tinyurl.com/692c5g>, or in automatic-translation English at <http://tinyurl.com/5fvbrp>).  
*¡Qué horror!*

**Jose M. Rojo** Departamento de Fisiopatología Molecular y Celular, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain  
e-mail: jmrojo@cib.csic.es

Contributions may be submitted to [correspondence@nature.com](mailto:correspondence@nature.com). Please see the Guide to Authors at <http://tinyurl.com/373jsv>. Published contributions are edited. Science publishing issues are featured at Nautilus (<http://blogs.nature.com/nautilus>), where we welcome debate.

# SUMMER BOOKS



## Grand tour with a geologist guide

A field guide to the formation of Italy's Apennine mountains mixes the science behind the dinosaur extinction with accounts of travel and Italian history, explains **Ted Nield**.

### **The Mountains of St Francis: Discovering the Geologic Events That Shaped Our Earth**

by Walter Alvarez

W. W. Norton: 2008. 288 pp.  
\$25.95, £16.99

In 1975, as Generalissimo Francisco Franco lay dying in Madrid, a party of geology students on an excursion to the mountains of northern Spain took a diversion to the Basque capital, Guernica. Had their leaders known a little more about art history, they would have realized that Pablo Picasso's iconic 1937 painting, the object of the detour, was still in exile in the United States at the time.

I was on that trip and remember this well-meaning yet misguided diversion because it was a rare event in my personal geological history. Sadly, when travelling the world at other people's expense, geologists will drive straight past every wonder of the world just to squeeze one more interesting roadside ditch into their itinerary. After reading *The Mountains of St Francis*, I feel certain that our trip would not have gone so awry had we been fortunate enough to have its author Walter Alvarez in our party.

If geologists built great cathedrals instead of ignoring them, they would probably put Alvarez's statue in a niche on the west front. Assisted by his father, Nobel prize-winning physicist Luis Alvarez, and nuclear chemists Frank Asaro and Helen Michel, Alvarez can claim priority as a principal discoverer of the global iridium anomaly that marks the junction between the Cretaceous and Palaeogene periods, the Cretaceous–Tertiary or KT boundary.

Iridium is rare on Earth's surface yet plentiful in space. The sudden spike in the element's abundance at the KT boundary on Earth is believed to have been caused by a comet or meteorite hit. The story of that discovery, and how it was linked to the dinosaur extinction and the offshore crater at Chicxulub, Mexico, was recounted in Alvarez's 1997 bestseller, *T. rex and the Crater of Doom* (Princeton University Press). It is this stunning series of discoveries with which Alvarez is most closely associated.

The particular roadside ditch in which Alvarez first discovered the iridium layer is not far from the town of Gubbio, near Perugia, in the Apennine Mountains in Italy. Alvarez has spent his career looking at the geology of

how that mountain range has been, and is still being, created. This forms the subject of his second foray into popular science.

Forming the backbone of Italy and enclosing the Adriatic embayment, the Apennines have long been difficult to reconcile with conventional plate tectonics alone. Their basic shape is governed by rock structures known as fold-thrust belts, where low-angle reverse faults in Earth's crust are formed as accumulated layers of sedimentary rock from former Adriatic seas have folded. The belts of rock become pushed up and over themselves to form upfolds, or ramp anticlines, at their leading edges.

Perplexingly, the compressional tectonics that push the crust together at the leading edges of these fold-thrust belts is followed by extensional tectonics behind, where the crust is pulled apart. This sequence of compression and extension calls for special explanation. It is believed to be caused by 'delamination and roll-back', whereby a slab of deep-lying continental crust undergoes mineral transformations that allow it to fall into Earth's mantle. This downward peeling process exerts drag on the crust above, thus creating compression at the leading

ILLUSTRATIONS BY P. HACKETT

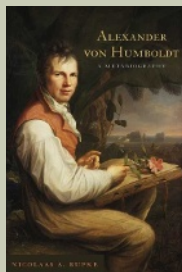
## NEW IN PAPERBACK

### **Alexander von Humboldt: A Metabiography**

by Nicolaas A. Rupke

(Univ. Chicago Press, \$24, £11)

Rupke looks at how Alexander Humboldt's life story has been adapted to suit changing philosophies or political ideals. "His metabiography helps us to appreciate the historical instability of any scientific life, not just one as complex as Humboldt's," wrote Steven Shapin (*Nature* **441**, 286; 2006).

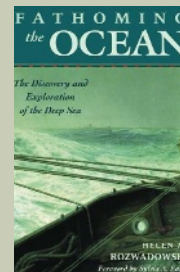


### **Fathoming the Ocean: The Discovery and Exploration of the Deep Sea**

by Helen M. Rozwadowski

(Harvard Univ. Press, \$19.95, £12.95)

This book explores the birth of deep-sea oceanography in the nineteenth century, covering the breakthroughs in gathering data and the social impacts. It explains how the presence of researchers on naval vessels led to cultural shifts for scientists, sailors and Western society.





edge and extension at the trailing edge.

Alvarez leads us to this hypothesis along a personal and scientific journey through many scientific, historical and cultural byways. The book reads like an extended field guide and notebook — Alvarez mingles the story of his involvement with Italian geologists with accounts of his travels, discoveries and what might be called an Italian history of the world, in which various historical figures receive proper credit for thinking of things first. Many of them, such as the seventeenth-century Danish geologist Nicolaus Steno, need no introduction to geological readers. But some Italian figures will be unfamiliar, such as the 'father of Italian geology' Giovanni Arduino, who gave us the now obsolete term 'tertiary' in the eighteenth century.

Alvarez obsesses a little about issues of priority, recalling my feeling that his other book was uncomfortably full of praise for those collaborators who had set their competing claims aside. Here, too, using many historical examples, Alvarez shows the reader how well scientists behave when science works.

I particularly relished the moral he draws from how alpine thrust faults were correctly explained. These faults are flat planes along which huge masses of rock have travelled horizontally, often for hundreds of kilometres. After years of doubt, geologists finally accepted that rocks could be pushed such large distances — their acceptance of plate tectonics sapped all controversy from the issue. Alvarez uses this to remind us of a phenomenon first identified by astronomers Alan Lightman and Owen Gingerich in 1992, whereby well-established ruling theories "develop a life of their own" and seem to take forever to collapse under the weight of conflicting evidence.

There is an irony here because the connection that Alvarez and others made between the iridium anomaly and the Chicxulub crater has become just such a ruling theory in recent years, one that sceptical scientists have challenged at their peril. One must conclude that the book was already with the publishers last year when Gerta Keller, a professor at Princeton University in New Jersey, persuasively debunked Chicxulub

as *T. rex*'s crater of doom by proving it was 300,000 years too old. Alvarez's frequent references to the crater as the uncontested 'smoking gun' would otherwise, presumably, have been phrased more circumspectly.

Like Alvarez's previous book, *The Mountains of St Francis* is a first-person participant history. This genre carries dangers, not least the pitfall of overplaying the author's centrality. Alvarez is given to grand gesture, and sometimes allows himself to come too close to writing what Stephen Jay Gould dubbed "cardboard history". For example, he avers that, by the 1930s, Alfred Wegener's ideas on continental drift had been "mostly rejected". True, Wegener's fellow geophysicists rejected his hypothesis because they thought it was physically impossible. But, as science historian Naomi Oreskes has shown, it was only in the United States — for a long time a bastion of old-fashioned continental fixism — that geologists rejected drift en masse. Elsewhere they remained more open-minded.

Alvarez deserves his place in posterity. Along

with others, including my late professor, Derek Ager, he has helped geologists to understand that uniformitarianism and gradualism are not the same thing. On the timescale of a planet, uniformity must also embrace rare, catastrophic events that may recur on timescales far beyond the duration of civilizations and even species.

Whatever the truth behind his sometimes overconfident side-statements about the end-Cretaceous extinction, meteorite impacts, the iridium anomaly, Chicxulub crater and science history, the tectonic story running through this book is compelling and engagingly told. It also holds appeal for lay readers, perhaps less so than his previous book, but those who venture into the mountains of St Francis with Alvarez will not regret it. I would make it required background reading for students of Earth science. It would certainly help counteract the occasional philistinism of those leading their field trips. ■

**Ted Nield** is editor of *Geoscientist* magazine and author of *Supercontinent*.

e-mail: ted.nield@geolsoc.org.uk

## Stalin's war on genetic science

### **The Murder of Nikolai Vavilov: The Story of Stalin's Persecution of One of the Great Scientists of the Twentieth Century**

by Peter Pringle

Simon and Schuster: 2008. 384 pp. \$26

It is not surprising, given the parlous state of Russia in the years following the Revolution, that its political system put ideology and practical outcomes above all else, including scientific fact. This was most evident in agriculture, where it was imperative to produce more food by whatever means. The consequences were tragic for the Russian people and for Nikolai Ivanovich Vavilov, Russia's greatest geneticist. Vavilov fell foul of Trofim Denisovich Lysenko who, through political manipulation and intrigue, came to dominate Soviet genetics. Peter Pringle's compelling book, *The Murder of Nikolai Vavilov*, tells the story of

the Lysenko affair with verve and pace. Pringle makes it clear how Vavilov's patriotism, dedication to science and determination to be open-minded led to his downfall and death.

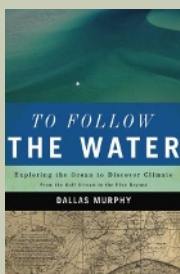
Vavilov was born in 1887 in Moscow into a comfortable, bourgeois family. In 1906 he entered the Petrovskaya Agriculture Academy, or Petrovskaya, one of many institutes established after the devastating famine of 1892. Russian agricultural practices lagged behind those of other European countries and the United States, and efforts to reform them were unsuccessful. Vavilov undertook "to work for the benefit of the poor, the enslaved class of my country, to raise their level of knowledge". This pledge, Pringle explains, drove Vavilov throughout his life.

After graduating, Vavilov spent a year researching wheat with Robert Regel at the Bureau of Applied Botany in St Petersburg,

### **To Follow the Water: Exploring the Ocean to Discover Climate**

by Dallas Murphy (Basic Books, \$15.95, £9.99)

As well as covering the history of human expansion across the globe and the science of oceanography, Murphy also gives first-hand accounts of life on a research vessel. "Meticulously following the waters of the Gulf Stream into the blue beyond, Murphy's book gets it right," wrote Arnold Gordon (*Nature* **449**, 407–408; 2007).

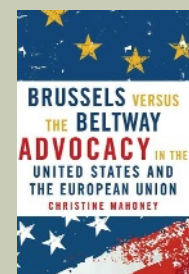


### **Brussels Versus the Beltway: Advocacy in the United States and the European Union**

by Christine Mahoney

(Georgetown Univ. Press, \$29.95)

Political decisions made in Washington DC and Brussels have global effects, but US and European advocacy styles are often assumed to be culturally different. This book challenges stereotypes, arguing that the context of issues and institutions is more important than differences between cultures.



before embarking on a two-year tour of European laboratories. His stay with William Bateson in Cambridge, UK, was the highlight. Bateson was the leading proponent of Gregor Mendel's work on inherited traits, rediscovered 10 years earlier, and wrote the first genetics textbook, *Mendel's Principles of Heredity*, published in 1909. Bateson's enthusiasm for Mendelian genetics seems to have rubbed off: Vavilov based his life's work on Mendelian principles and their elaboration by, among others, fly geneticist Thomas Hunt Morgan. Bateson had led an expedition to the Russian Steppes in 1886 to examine the interactions of environment and species variability. Pringle suggests that this may have inspired Vavilov to undertake similar expeditions to search for crop varieties whose traits made them suitable for particular environments, such as dry or cold regions.

On Vavilov's return to Russia and the Petrovskaya, he was sent to investigate why soldiers on the Persian front were falling ill after eating bread. Vavilov used the assignment to collect varieties of plants growing in the harsh climate of the Pamir mountains, in the hope that these hardy plants might be cultivated in northern Russia to provide more food for the Soviet people. Vavilov endured great hardship in travelling to such remote regions, trips that would now be unthinkable without insulated jackets, mobile phones and satellite navigation.

Vavilov returned from the Pamirs in 1916 to find Russia in political turmoil. In March 1917, Tsar Nicholas II abdicated and by October, the Bolsheviks had seized the reins of government, plunging the country into civil war. Nevertheless, Vavilov's career began auspiciously — he took up a full professorship at the University of Saratov, a large city on the Volga river some 700 kilometres southeast of Moscow. Vavilov mounted expeditions to Afghanistan, Ethiopia, Eritrea, North and South America and the Mediterranean, seeking plants that might increase agricultural productivity in Russia. He regarded this as an essential task after the disastrous collapse of Soviet agriculture that followed the consolidation of land and labour known as collectivization that began around 1929. Vavilov's



collection of 250,000 seeds of cultivated plants and their varieties was the most extensive in the world. In 1930, he was appointed director of the Institute of Genetics of the USSR Academy of Sciences in recognition of his position as the country's leading plant geneticist and his international reputation. Just six years later, Vavilov was in disgrace.

His nemesis Lysenko was born in 1898 into a peasant family. Unusually for the time, he attended a school of agriculture and horticulture; clever and ambitious, he aspired to make great contributions to Soviet science. His big break as an agricultural researcher came in 1927, when the newspaper *Pravda* reported his work on changing the time of sprouting in seeds by exposing them to differing periods of cold temperatures, known as vernalization. The reporter noted that Lysenko was working for the people, not carrying out research for its own sake by studying the "hairy legs of flies". Lysenko promoted himself as the discoverer of vernalization, although it had been known since 1858, and trumpeted it as a solution to the Soviet Union's chronic food shortages.

Lysenko claimed that plants could be 'educated' so that the changed germination time became heritable after several generations of vernalization. This was a variant of Lamarckism, or the inheritance of acquired characters, that had been discredited first by August Weismann's distinction between germ cells and

somatic cells, and second by Mendel's work. Scientists rejected Lysenko's claims, but by skilful manipulation of the political situation throughout his career, Lysenko scaled the Soviet scientific hierarchy. He was twice awarded the Order of Lenin, and became president of the Lenin Academy of Agricultural Sciences of the USSR, a full member of the country's Academy of Sciences and a member of the Supreme Soviet.

The conflict between Lysenko and the 'Mendelian-Morganists' came to a head in 1936 at a conference at the Lenin Academy. Despite geneticists' devastating scientific critique of Lysenko's claims, the government-controlled press declared Lysenko the winner. Attacks on Vavilov's position increased and Lysenko consolidated his position. Senior scientists in the Soviet administration were among the victims of Stalin's Great Purge, when perhaps as many as one million anti-revolutionaries and enemies of the people were executed over two years, including Muralov, president of the Lenin Academy. Lysenko took his place to become Vavilov's boss. In October 1939, the Central Committee of the Communist Party of the Soviet Union held another genetics conference. This again ended in triumph for Lysenko.

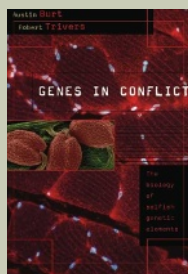
Why were the reins of Soviet agriculture held by a charlatan whose policies were disastrous? As Pringle makes clear, Lysenko prospered because he promised rapid advances in agriculture that were seized on by a Soviet government desperate to feed thousands of citizens dying of starvation. Lysenko promised Stalin that new strains of wheat and other crops with desirable traits could be produced within 3 years, much quicker than the 12 years that Vavilov required. Perhaps as importantly, Lysenko's views of genetics were in sympathy with prevailing Marxist dogma.

Experts, by virtue of their education and role, were members of the bourgeoisie and regarded with suspicion in Russia. There was a strong political movement to replace the intelligentsia with elevated peasants and other members of the proletariat, even if they were untrained and ill-fitted to their new posts. Lysenko was one such example. Vavilov, by contrast, was an educated, well-travelled businessman's son who was

### Genes in Conflict: The Biology of Selfish Genetic Elements

by Austin Burt and Robert Trivers  
(Harvard Univ. Press, \$24.95, £16.95)

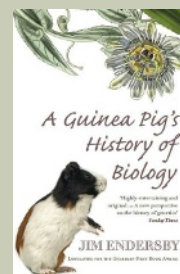
The book examines how selfish genes thrive despite damaging their carrier. Burt and Trivers describe how such genes spread through populations by distorting transmission on a molecular level and increasing their own replication, or by changing how an organism treats its kin.



### A Guinea Pig's History of Biology: The Plants and Animals Who Taught Us the Facts of Life

by Jim Endersby (Arrow, £8.99)

From Mendel's peas to modern laboratories, Endersby's entertaining book explores the history of science through the plants and animals that provided answers or, often, more questions. Covering the familiar ground of mice and fruitflies, the book also highlights organisms that led scientists up the wrong path, such as the zebra-like quagga.





thought to be susceptible to foreign influences.

And why did Vavilov not fight Lysenko earlier and more aggressively? Pringle demonstrates that Vavilov was guided by his student pledge to help the Soviet people and that he was committed to exploring all leads, however improbable, that might increase food production. Vavilov encouraged many scientists, including Lysenko, to test different approaches. Naively, Vavilov did not expect that Lysenko would play by political rather than scientific rules. At a 1948 session of the Lenin Academy, Stalin was so determined that Lysenko should triumph that he drafted Lysenko's opening remarks himself, emphasizing the correctness of Lamarckian thinking. A letter included in the official report ended: "Glory to the great Stalin ... coryphaeus of progressive science!"

Neither Vavilov nor his work featured in this session. Following the 1939 conference, Lysenko had progressively dismantled Vavilov's institute, but Vavilov had remained free even as criticism of him became ever more vituperative. Then, on 6 August 1940, while collecting plants in the Ukraine, Vavilov was seized by the Soviet secret police and taken to Moscow. Pringle's account of Vavilov's 11-month interrogation is horrifying. In July 1941, Vavilov and two colleagues were tried and sentenced to death. Vavilov's appeal to the Presidium of the Supreme Soviet was turned down, but a personal plea to the head of the secret police led to his sentence being commuted to life imprisonment. His colleagues were shot. Vavilov died of starvation on 26 January 1943 in a prison in Saratov, the city where he had begun his illustrious career 26 years before.

Even now, politics continues to trump good science, as is evident from the delays in reducing global carbon emissions. Pringle's very readable account is a timely reminder that public policies must be based on rational decisions drawn from the best data available. ■

**Jan Witkowski** is executive director of the Banbury Center and professor in the Watson School of Biological Sciences at Cold Spring Harbor Laboratory, New York 11724, USA. He is co-author of *Recombinant DNA: Genes and Genomes*. e-mail: witkowski@cshl.edu

## Betting on black holes

**The Black Hole War: My Battle with Stephen Hawking to Make the World Safe for Quantum Mechanics**

by Leonard Susskind

Little Brown: 2008. 416 pp. \$27.99

The idea of a black hole in space produced by the collapse of a massive star dates back to the 1930s, but it was only in the 1960s that astrophysicists began to understand the details. Evidence now suggests that black holes do exist, and are a key element of the great cosmic story. Yet some of their predicted properties remain puzzling and threaten cherished physical laws.

One long-running conflict concerns the fate of the material that implodes to form a black hole. In his new book *The Black Hole War*, theoretical physicist Leonard Susskind of Stanford University, California, describes how he sparred with Stephen Hawking of the University of Cambridge, UK, about this thorny issue.

Black holes earn their name because their gravity is so strong they trap even light, appearing black from the outside. According to Einstein's general theory of relativity, the ball of matter that implodes to create the hole continues to shrink, meeting an uncertain fate at the centre and leaving behind it a region of empty space. Because physicists believe nothing can go faster than light, no information or material should escape the hole. Practically all the information about the collapsed star would be lost from the outside Universe, making it impossible to tell whether the star was made of matter, antimatter or green cheese; once inside the hole, its external physical properties would be the same.

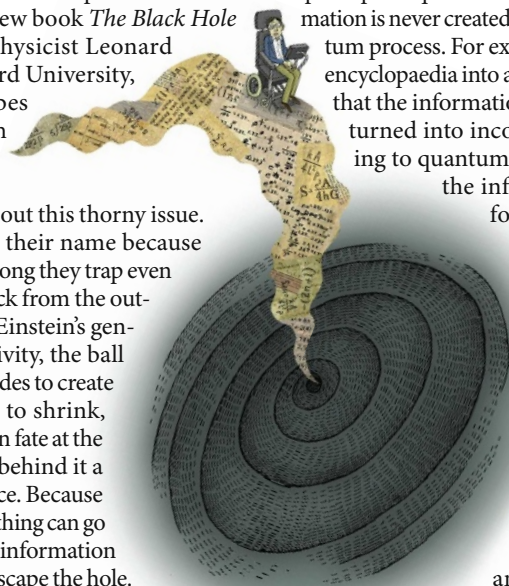
A twist arose in 1975 when Hawking argued that black holes are not perfectly black. By

applying quantum mechanics to the formation process of a black hole, he predicted that it must radiate heat. Because heat is a form of energy, this radiance would have to be paid for by gradually reducing the hole's mass, because energy and mass are equivalent in general relativity. In time, the hole would evaporate away completely, leaving behind just heat energy, mostly in the form of photons, with a few other particles such as neutrinos and electrons.

On an immense timescale, vastly longer than the current age of the Universe, the black hole would convert the entire star into heat. Hawking concluded that the heat energy would look the same whatever the star was made of originally. Many physicists did not like this conclusion. A principle of quantum mechanics is that information is never created or destroyed in a quantum process. For example, if you throw an encyclopaedia into a furnace, it might seem that the information is irretrievably lost, turned into incoherent heat. According to quantum mechanics, however, the information is lost only for practical purposes.

The infrared photons emitted by the embers still contain the original information, but in a hopelessly scrambled form that renders it inaccessible to us. Leading particle physicists, led mainly by Susskind and Gerard 't Hooft of the University of Utrecht in the Netherlands, declared that the same would be true of black holes — the information about the original star would be enfolded in the emitted heat. It might be jumbled up, but it should all be there.

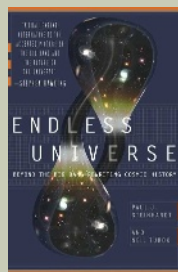
Hawking begged to differ. I recall him going through the argument during a conversation in



### Endless Universe: Beyond the Big Bang — Rewriting Cosmic History

by Paul J. Steinhardt and Neil Turok  
(Broadway, \$14.95, £8.99)

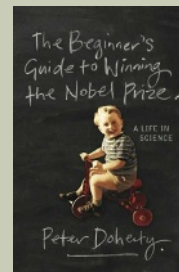
Two theoretical physicists challenge the widely accepted Big Bang Theory. Turok and Steinhardt explain how new developments in astronomy, cosmology and physics have led them to develop their own 'Cyclical Universe' theory to explain cosmic origins.



### The Beginner's Guide to Winning The Nobel Prize: A Life in Science

by Peter Doherty (Columbia Univ. Press, £10.95)

Doherty's highly readable book describes how to succeed in science and hopefully win that coveted prize. "For the aspiring young scientist, or a student considering a scientific career, Doherty opens the vault to the world of science, explaining how it works and how to get on," according to Peter Parham's review (*Nature* **443**, 755–756; 2006).



1978. He reasoned that crucial bits of information about subatomic particles cannot be carried by photons, so other types of particles such as electrons or quarks would also be needed. But there are not enough of these other particles in the radiation to embody all the information, because most of the heat energy from a black hole is in the form of photons. The conundrum became known as the information paradox. It may seem like theoretical nit-picking, but the paradox goes to the very heart of physical theory and its description of reality.

The momentous conclusion that a black hole swallows and permanently obliterates physical information didn't bother Hawking, whose background was in gravitational theory and space-time geometry rather than particle and quantum physics. With theorist Roger Penrose of the University of Oxford, UK, he proved that space and time could have boundaries or edges, called singularities, at which information might enter or leave the Universe. The general theory of relativity predicts that such a singularity lurks at the centre of a black hole, where the gravitational field and space-time warp become infinite. As a consequence, the imploding star's information might disappear from space and time through the hole's singularity. Hawking was sufficiently confident to place a bet with theoretical physicist John Preskill at the California Institute of Technology in Pasadena.

An uncertainty in Hawking's argument was the fate of the singularity. If the black hole evaporated completely then the singularity must presumably disappear too, but the details depend on an incomplete field of physics known as quantum gravity. In recent years, work in string theory — to which Susskind has made seminal contributions — has placed quantum gravity on a more secure foundation. Armed with such arguments, Susskind and others gathered support for the position that information is conserved, against Hawking's claim. In 2004, Hawking called a press conference and announced to the world that he had changed his mind. Black holes did not, after all, irreversibly annihilate information, he said. The bet with Preskill was duly settled in the form, appropriately enough, of an encyclopaedia.

*The Black Hole War* charts the ups and downs of this lengthy yet good-natured dispute. Susskind skilfully explains the subtleties of the physics that underlie the issue, and includes anecdotes to enliven the technical details. He has stuck to his guns for many years, but nevertheless resists the temptation to gloat over Hawking's eventual capitulation.

Is the matter laid to rest? I don't think so. Hawking justified his reversal by sketching out a calculation, but quantum gravity is still too unrefined for a rigorous proof. The weak point is that, in quantum gravity, the singularity can be replaced by a space-time region with a complicated and changing topology, allowing information to shift from one region of space-

time into another disconnected one, perhaps from our Universe to a newly born 'baby universe'. Susskind dismisses this possibility, but the matter is far from resolved. It may be that if we consider the entire 'meta-verse' of all spatial regions, information is never lost. But if we restrict attention to a single universe, or connected region of space, then information can in fact leak out. Deciding the matter is a task for a future generation of theoretical physicists. ■

**Paul Davies** is director of the Beyond Center for Fundamental Concepts in Science at Arizona State University, Tempe, Arizona 85287-6505, USA. His latest book is *The Goldilocks Enigma: Why Is the Universe Just Right for Life?* e-mail: paul.davies@asu.edu

## Maths and mad hatters

### Lewis Carroll in Numberland: His Fantastical Mathematical Logical Life

by Robin Wilson

Allen Lane/Norton: 2008. 237 pp/208 pp.  
£16.99/\$24.95

Legend has it that Queen Victoria was so enchanted by *Alice's Adventures in Wonderland* that she insisted on Lewis Carroll's next work being sent to her. One can imagine her expression as she opened the book that arrived, entitled *An Elementary Treatise on Determinants*.

Charles Lutwidge Dodgson had many careers. He is best remembered for the sublime nonsense verse he wrote under the name Lewis Carroll. He was a pioneering children's photographer and a lay clergyman admired for his sermons. Before all else he was a mathematician who taught generations of students at the University of Oxford, UK, contributed to the fields of geometry, algebra and logic, and used games and puzzles to entertain and instruct. In *Lewis Carroll in Numberland*, mathematician Robin Wilson reveals Dodgson to be the grandfather of recreational mathematics.

He was precocious, orthodox and craved variety. Born in 1832 in Cheshire, UK, Dodgson

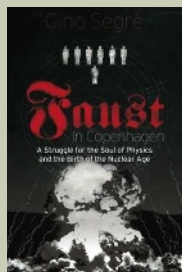
was a lecturer at Oxford by his early twenties. At a time when non-Euclidean geometries were catching on, he wrote a four-act play stubbornly arguing that Euclid should remain at the centre of the Oxford curriculum. He invented a method to find the determinants of large matrices, but his strange notation meant that it never caught on. Later, he sought mathematical remedies for real injustices, suggesting tie-break methods for parliamentary elections to his friend Lord Salisbury, and devising a way to make lawn tennis tournaments fairer to the runners-up.

Some work was ahead of its time, especially his efforts to bring mathematics to young people. Although pupils complained of his "singularly dry and perfunctory manner" in the classroom, Dodgson's gift for teaching shone through in dozens of self-published guides for students, and in his letters to children. Wilson shows that he found humour in the plainest of subjects and did not underestimate his young correspondents, once commenting that intelligence seemed to vary inversely with size. In person, he drew their attention using guessing games and feats of memory. He could recite the first 71 digits of pi using a series of nonsense couplets as memory aids, and once contrived

### Faust in Copenhagen: A Struggle for the Soul of Physics and the Birth of the Nuclear Age

by Gino Segrè (Pimlico, £12.99)

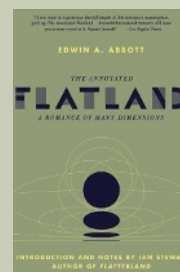
Segrè relays the 1932 gathering of seven leading physicists that "symbolized the end of the political neutrality of physics and physicists, coinciding as it did with the arrival of Hitler and crucial discoveries in nuclear physics that would make possible the subsequent development of the atomic bomb", wrote Finn Aaserud (*Nature* **448**, 869–870; 2007).



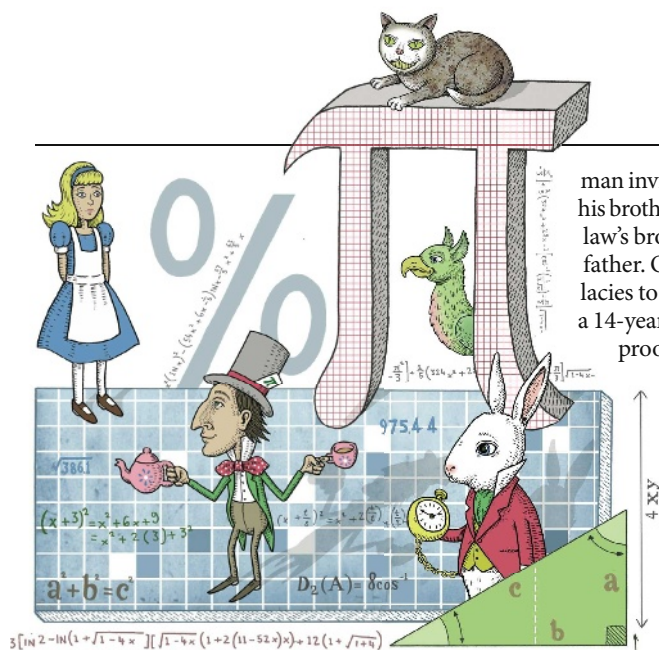
### The Annotated Flatland: A Romance of Many Dimensions

by Edwin A. Abbott. Introduction and notes by Ian Stewart (Perseus, \$17.95, £10.99)

Abbott's satirical tale of A. Square's journey to The Land of Three Dimensions is annotated by mathematician Ian Stewart, providing helpful background on this maths fantasy. Abbott's ideas were thought radical by the Victorians but are now commonplace in both science and science fiction.







man invited his father's brother-in-law, his brother's father-in-law, his father-in-law's brother, and his brother-in-law's father. Others were in the form of fallacies to debunk. Dodgson once asked a 14-year-old boy to find the flaw in his proof that  $2 + 2 = 5$ , which Wilson reveals to be a stealthy division by zero. A few problems hinged simply on a pun.

Later in life, Dodgson taught symbolic logic with a board game that used red and grey counters on a set of nested squares, which he believed superseded the overlapping circles championed by British logician John

symmetrical poem that can be read vertically and horizontally.

*Lewis Carroll in Numberland* is not a conventional biography. Robin Wilson has winnowed Dodgson's prodigious output into a first-rate scrapbook of proofs and puzzles. Sadly, his tone is often fawning and flat — not up to the standard of mathematical storytelling he set in his previous book, *Four Colours Suffice* (Allen Lane, 2002), on the history of the conjecture that four colours can fill any map without any bordering countries sharing a colour. By immersing us in Dodgson's correspondence, however, Wilson conjures the spirit of a man who delighted in paradox yet insisted on precision, who held fiercely to the ancients while straining to understand the world around him, and who wanted most of all to stump everyone he knew. Writing for work or pleasure, for children or adults, Wilson shows that Dodgson turned the most sober of problems into child's play.

"Some perhaps may blame me for mixing together things grave and gay," he wrote as Lewis Carroll in an insert to his nonsense poem *The Hunting of the Snark*. But, he continued, "I do not believe God means us thus to divide life into two halves."

**Jascha Hoffman** is a writer based in New York. e-mail: [jascha@jaschahoffman.com](mailto:jascha@jaschahoffman.com)

an algorithm that could give the date of every Easter Sunday until 2499.

Sooner or later every child who knew Dodgson would receive a brain teaser. Published in collections with titles such as *A Tangled Tale* and *Pillow Problems Thought Out During Sleepless Nights*, many of these word problems required the dutiful application of algebra, trigonometry or geometry. Some needed mere patience and common sense. One devious puzzle asked how many guests would come to a dinner party if a

Venn. In Dodgson's *Game of Logic*, published under his pen name to gain a wider audience, one can see some of the punctilious lunacy of the Mad Hatter. Following chains of inference he called 'sillygisms', he led readers from reasonable premises to conclusions such as "Babies cannot manage crocodiles", "No banker fails to shun hyaenas" and "No bird in this aviary lives on mince-pies". These examples are perhaps less interesting as logic than as the stirrings of a systematic kind of literature, also apparent in his

## The creationist controversy

### Only a Theory: Evolution and the Battle for America's Soul

by Kenneth R. Miller

Viking: 2008. 256 pp. \$25.95

The United States has a big problem: although we maintain a strong scientific establishment, competitive with the rest of the world in many fields, we also have some of the most backwards proponents of superstitious nonsense in both our electorate and at the highest levels of politics. It is an embarrassment to host laboratories that are at the forefront of scientific research in the same country where presidential

candidates are discussing whether Earth is really 6,000 years old as some Bible scholars say, or whether they believe in evolution.

Science and evolution have an advocate in Kenneth Miller, one of North America's eminent knights-errant, a scientist who is active in defending evolutionary theory in the conflict between evolution and creationism. He has been at the centre of many recent debates about science education, most prominently testifying against intelligent design creationism in Pennsylvania's Dover trial, which decided that intelligent design was a religious concept that should not be taught in public

schools. He is also a popular speaker, offering the public a grass-roots defence of good science education. Miller's new book *Only a Theory* is a tour of creationist misconceptions about evolution, such as the one referred to in the book's subtitle — a creationist predicted an inevitable victory in the Dover trial because evolution is "only a theory". The book is also a celebration of the power of evolutionary theory to explain our existence.

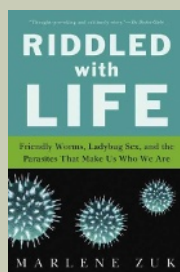
Miller is a fine writer who sharply addresses the details of the arguments about intelligent design creationism. When tackling old chestnuts such as the "only a theory" complaint, or Michael Behe's argument for a maximum limit for the number of genetic mutations, or William Dembski's rehash of William Paley's

### Riddled with Life: Friendly Worms, Ladybug Sex, and the Parasites That Make Us Who We Are

by Marlene Zuk

(Harvest, \$14, £8.99)

An evolutionary biologist enthusiastically argues that parasites are a driving force behind evolution and that their effects still mould us today. Parasites have shaped us physically and culturally, and affect our minds on a daily basis.

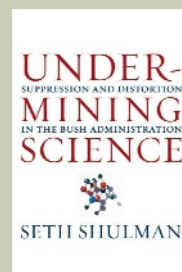


### Undermining Science: Suppression and Distortion in the Bush Administration

by Seth Shulman

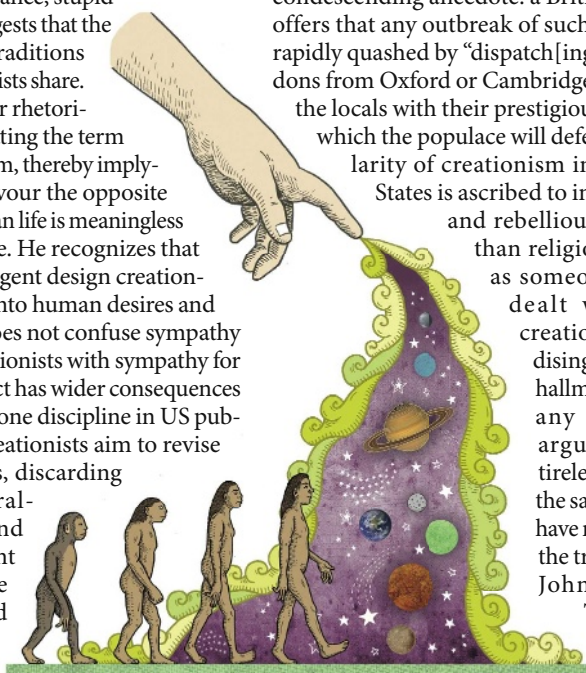
(Univ. California Press, \$16.95, £9.95)

Shulman explores how US politicians and scientists have misrepresented science to push their own agendas. "A concise, straightforward case history of the politicization of science, ideal for courses on the history, philosophy, sociology and ethics of science," wrote John Horgan (*Nature* **445**, 365–366, 2007).



watchmaker argument for complexity, Miller discusses the contemporary biological explanations while refuting the errors.

Miller is sympathetic to the creationists' perspective but opposes them uncompromisingly. The book does not try to place the blame for creationism on ignorance, stupidity or malice, but suggests that the ideas are rooted in traditions and values that biologists share. He admires the clever rhetorical trick of appropriating the term 'design' for creationism, thereby implying that scientists favour the opposite and believe that human life is meaningless and without purpose. He recognizes that the concept of intelligent design creationism taps effectively into human desires and prejudices. Miller does not confuse sympathy for the intent of creationists with sympathy for its effects. The conflict has wider consequences than the teaching of one discipline in US public schools — the creationists aim to revise what science means, discarding rationalism, naturalism, materialism and other Enlightenment values to incorporate the supernatural and loosen the rigour of all sciences.



*Only a Theory* deals poorly with one central aspect of this battle: why this problem is so much greater in the United States than elsewhere. Miller's rationalizations are sometimes painful to read. Europe's relative freedom from the scourge of creationism is explained with a condescending anecdote: a British colleague offers that any outbreak of such nonsense is rapidly quashed by "dispatch[ing] a couple of dons from Oxford or Cambridge" to overawe the locals with their prestigious degrees, to which the populace will defer. The popularity of creationism in the United States is ascribed to independence and rebelliousness rather than religiosity, which, as someone who has dealt with many creationists, I find disingenuous. The hallmark of almost any creationist argument is the tireless bleating of the same points we have rebutted since the trial of teacher John Scopes in Tennessee in 1925, which tested the

law on teaching Darwinian evolutionary theory; the only twists come from new creationist authorities that enter the fray. An equivalent US variant of Miller's British anecdote is that the enemies of science need only dispatch Dembski or Behe from the Discovery Institute in Seattle, Washington, to stir up more doctrinaire creationism among school boards and in elections and churches. To call US citizens more independent-minded than European citizens flatters the creationists too much and demeans Europeans.

If Miller is on shaky ground in his explanations of the origins of creationism, he is rock-solid on where the creationists want to take us: "To the intelligent design movement, the rationalism of the Age of Enlightenment, which gave rise to science as we know it, is the true enemy ... science will be first redefined, and then the 'bankrupt ideologies' of scientific rationalism can be overthrown once and for all." Although his own religious leanings blind him to conflict between faith and science, they also give him insight into both sides of the struggle. *Only a Theory* is a useful overview of a perilous political attack on the nature of science. ■

**PZ Myers** is associate professor of biology at the University of Minnesota Morris, 600 East 4th Street, Morris, Minnesota 56267, USA, and author of the blog Pharyngula. e-mail: pzmyers@gmail.com

## Fictional quantum conspiracy

### Final Theory: A Novel

by Mark Alpert

Simon & Schuster: 2008. 368 pp.  
£12.99, \$24

Most conspiracy theorists focus on political cover-ups. But science is an excellent catalyst for this sort of paranoia too: so entrenched is the stereotype of the mad researcher that it is not surprising people might suspect that someone, somewhere, is hiding something for nefarious gain. Physics in particular lends itself to these

sorts of fears. Whereas most people can conceptualize tangible sciences such as biology, the quantum world is, by its very nature, largely ungraspable and seems to simmer with deadlier force. The Manhattan Project, which still casts a long and sinister shadow in the popular imagination, certainly didn't improve its reputation.

Mark Alpert's debut novel *Final Theory* is classic conspiracy fodder. It posits that Albert Einstein, who in real life spent his last decades failing to unify quantum mechanics with relativity, instead succeeded. Realizing the military

implications, yet incapable of destroying the beautiful mathematical proofs outright, Alpert's fictionalized Einstein decides to bury the information until mankind has outgrown its warlike ways. He duly entrusts various pieces of the puzzle to a select group of young protégés. Many years later word leaks out, and soon the US government and a rogue terrorist are hot on the trail. One by one, the protégés — now old men — are tortured and killed.

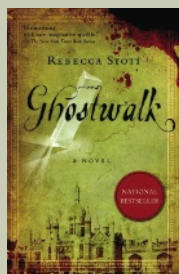
David Swift, the protagonist, is a failed physicist-turned-science historian who witnesses the dying words of one of these men, his former PhD supervisor. Before long, Swift has been taken prisoner by the FBI and, after escaping,

### Ghostwalk: A Novel

by Rebecca Stott

(Spiegel & Grau/Phoenix, \$14.95/£7.99)

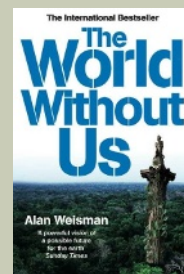
Seventeenth-century murder and present-day mysteries become entangled in Stott's *Ghostwalk*. Fine historical research is combined with a modern literary thriller when Lydia Brooke is asked by a former lover to complete his late mother's great work — a history of Isaac Newton's involvement in alchemy.



### The World Without Us

by Alan Weisman (Virgin, £8.99)

If humans disappeared, what would the world look like? Using evidence from places where war or disaster have created no-go zones for humanity, Weisman describes what would happen in our absence and what would be left behind. Stuart Pimm wrote: "There is no guarantee that even with all the pieces, we would be able to put nature back together again," (*Nature* **448**, 135-136; 2007).

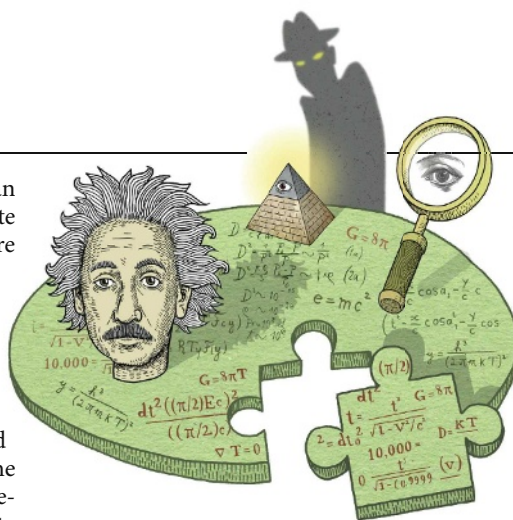




goes on the run with Monique Reynolds, an up-and-coming string theorist, in a desperate attempt to find the secret information before it falls into the wrong hands.

The book is reminiscent of Dan Brown's *Angels and Demons*, about the exploitation of antimatter. But this is no ordinary thriller. First, Alpert can actually write. Like many 'lab lit' authors, he is clever with scientific metaphors: Monique is at one point described as "unyielding and unstoppable, bending the whole fabric of spacetime around her". Moreover, Alpert has made an effort to integrate serious physics into the plot. As a former grad student turned Einstein biographer, Swift knows the great man's work intimately and can explain the basics on behalf of the reader. In the author's note, we learn that Alpert has a lot in common with his protagonist. Both share a physics education, a defection to a peripheral career (Alpert currently writes for *Scientific American*) and, like Swift in the novel, Alpert is author of a seminal research paper that is enjoying a re-examination.

Alpert manages to avoid some of the usual fictional-scientist stereotypes. His coup is Reynolds, a black female physicist who drives



a Corvette with the number plate 'STRINGS'. Her geek chic is as far from the boffin cliché as you can get; indeed it is a reasonable representation of what modern scientists can be like. Alpert does occasionally slip: grad students are described as pale, gangly, poorly dressed and bespectacled, and the last Einstein protégé left standing goes mad while seeking to exploit the theory for his own ends. Yet right up until the point he starts waving around an Uzi, the protégé's 'madness' is relatively harmless, fixated on unlimited energy and new medicines.

It is disappointing when these wild but admirable dreams degenerate into frank evil.

It would have been more elegant had Alpert explored scientists' obsessive nature without actually crossing that line. Nevertheless, even this character is not half as mad as the rogue terrorist nor as evil as the FBI. The flip-side of yearning for plausible scientific characters in fiction is to recognize that, as human beings, scientists should be allowed to be as prone to crazy or bad behaviour as any other member of society.

The more disturbing stereotypical trait in the book, however, is that scientists shouldn't meddle with things they aren't meant to know. Swift "thought he could get a glimpse of the Theory of Everything without suffering any consequences, and now he was being punished for this sin of pride, this rash attempt to read the mind of God". This could be a sentiment straight from *The Clouds*, Aristophanes' cautionary comedy about the hubris of the sophist school, or the myth of Icarus, who flew too close to the Sun. Haven't we moved on a bit since then? ■

**Jennifer Rohn** is a cell biologist at University College London, Gower Street, London WC1E 6BT, UK, and editor of [www.lablit.com](http://www.lablit.com). Her novel *Experimental Heart* will be published later this year. e-mail: [jenny@lablit.com](mailto:jenny@lablit.com)

## Inside the mind of a marathon runner

### What I Talk About When I Talk About Running

by Haruki Murakami

Translated by Philip Gabriel

Harvill Secker/Knopf: 2008. 192 pp.

£9.99/\$21

Reading, writing and running: three skills I did not expect to encounter alongside each other with much passion. I grew up playing sports and desired a physically active career. Of the three skills, running came to me last and the hardest. I took it up after the Athens 2004 Olympic Games. One Olympiad later, I find I share these interests with Japanese novelist Haruki Murakami, who has written

a memoir about the role that marathon running plays in his life.

The Athens Olympics was a turning point in my amateur athletics career because the city's heavy smog made me rethink my asthma treatment. I enjoyed anaerobic or short-burst events, but quickly became short of breath. After taking my fix of salbutamol — technically doping, if not prescribed — I could continue in some limited fashion, but endurance events eluded me. I decided on returning from Athens to start taking a preventive inhaler, beclomethasone dipropionate. My doctor had prescribed it, but I had never taken it, objecting to being permanently medicated for a mild and reasonably well-controlled condition. Within a

month of using it, I could run for an hour without taking a deep breath or additional medication. The experience was transformative.

This is how it felt reading *What I Talk About When I Talk About Running*. At the start, I thought Murakami and I were different: he is human and I am a cyborg. When I run, I am motivated by the thought that this should not be possible, that I am defying nature. I feel 'better than well'. Murakami started running when he was 33, about the same age I am now, so I hoped to find some common ground. And so it proved.

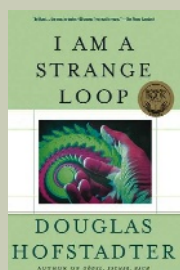
I was asked to review Murakami's memoir in the context of my expertise on the ethics of biotechnological enhancements. Perhaps

### I Am A Strange Loop

by Douglas R. Hofstadter

(Basic Books, \$16.95, £9.99)

Hofstadter extols his views on the nature of consciousness and the self. The book provides an interesting journey whether you agree with his conclusions or not. He "whisks us away to tangle with ever more layers of paradox and wonderfully mind-wrenching questions," wrote Susan Blackmore (*Nature* **447**, 29–30; 2007).

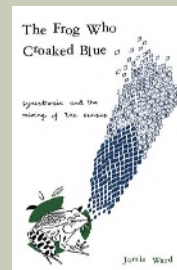


### The Frog who Croaked Blue: Synesthesia and the Mixing of the Senses

by Jamie Ward

(Routledge, \$16.95, £8.99)

A fascinating introduction to synaesthesia, explaining how the trait gives insight into the way the senses are organized. Ward also delves into other sensory experiences, such as phantom limbs and sympathetic touch, in an accessible introduction to this growing research field.





this text might persuade me of the value of remaining unenhanced by technology. After all, the experience of a long-distance runner — international novelist or not — is typically existential, with narrative usually more important than competition to an individual's performance. Runners refer to being in the 'zone' or the 'runner's high', the latter describing the euphoria experienced when running long distances. What struck me about Murakami was his ability to reconcile the two dimensions of competitiveness and personal narrative.

Murakami speaks of his body in mechanical, performative terms, attributing autonomy to each body part, thereby invoking the prospect of intelligent biology. His muscles talk to him, plead with him and sometimes work with him. His relationship with various parts of his body is an exemplar of Cartesian mind-body dualism, and accords with the science of long-distance running more generally. Murakami's

process of self-dialogue, reflected in his title, is a feature of a long-distance runner's introspection, a motivating mantra without which the perception of pain and awareness of the practice's futility might return.

Murakami's primary mode of performance enhancement is training, and critics and advocates of the integration of humans and technology should pause to reflect on that. Amid the mire of moral discourse on enhancements such as designer steroids, competitive sport comprises a technological relationship between biology and artifice. Murakami celebrates technological support in various, non-doping ways. For example, he coats his body in Vaseline before donning a wetsuit at the start of a triathlon to improve the efficiency of his switchover from swimming to cycling, when he must remove the suit. He uses a feather-light bicycle to optimize his speed and explains how competitive cycling is unlike riding for leisure: generating power on the up-pedal motion

changes the muscle group required.

His running shoes are light and well padded, but not enough to prevent the knee damage he encountered after his 62-mile ultramarathon in Japan. Murakami is at ease with the prospect of the long-term damage that accrues from his running, speaking of it as an inevitability of the pursuit. He does not dwell on the discomfort: "Pain is inevitable. Suffering is optional."

Running nourishes Murakami's writing. It allows him space, a void that he sees as a necessary encounter with nothingness to balance the verbose side of his life. Like many mountaineers who are compelled to climb, he runs because some unknown force makes him do so — his legs need to run, he says.

Readers who hope to understand Murakami through this memoir might feel unsatisfied. It is not about Murakami's life overall, but of his life as a runner. No great mysteries are uncovered about his writing, where it comes from, what inspires it, or what his books mean to him. The style is different from his other literary works. Yet it reveals that, for Murakami, books and running cannot be separated. With this his only published memoir, it is all we have of the writer at his most intimate.

*What I Talk About When I Talk About Running* reveals what kind of man Murakami is, rather than describing what he has done. He is a humble, self-effacing author who struggles with the idea of writing about himself, finding his fortunes to be a matter of unlikely chance. He succeeds because he has a strong sense of his own identity, his goals and expectations.

After one ultramarathon, around 25 full marathons, countless half-marathons and triathlons, and with his prodigious literary success, Murakami is a great example of the view that exercise stimulates the mind. I still wouldn't run a city marathon during any summer, certainly not in Athens or Beijing, without some kind of technological enhancement. ■

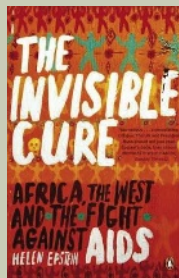
**Andy Miah** is reader in new media and bioethics at the University of the West of Scotland, Ayr KA8 0SR, UK. He is author of *Genetically Modified Athletes* and editor of *Human Futures: Art in an Age of Uncertainty*.

e-mail: email@andymiah.net

### The Invisible Cure: Africa, the West and the Fight Against AIDS

by Helen Epstein (Penguin, £9.99)

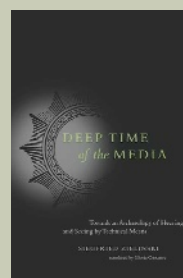
Challenging orthodox wisdom, "Epstein combines personal research and corroborative evidence from others to posit the view that where Africa's AIDS rates are highest, the key difference is not the numbers of sexual partners, but the timing", wrote Stephen Lewis and Paula Donovan (*Nature* **447**, 531-532; 2007).



### Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means

by Siegfried Zielinski (MIT Press, \$19.95, £12.95)

Focusing on how devices for hearing and seeing connect audience and creator, Zielinski highlights models and machines that changed the face of the media landscape, many overlooked by historians. These show that simple tools did not predictably lead to complex machinery.





## NEWS &amp; VIEWS

## NANOTECHNOLOGY

# Patterns from molecular corrals

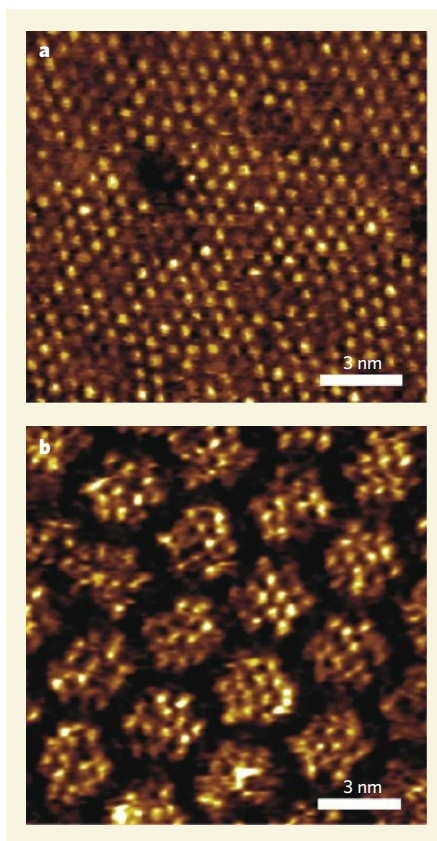
Michael Grunze

**Many nanotechnology devices will require components that consist of arrays of molecules positioned on surfaces with nanometre precision. One way to make these is to let the molecules organize themselves.**

A major challenge in nanotechnology is to find a way of positioning molecules and atoms on surfaces in regular patterns, with nanometre precision yet over large surface areas. Reporting on page 618 of this issue, Madueno *et al.*<sup>1</sup> describe just such a method. They have corralled self-assembling films within the pores of a two-dimensional network of molecules on a substrate, thereby forming patches of film in a repeating pattern that extends over a wide area. Each patch consists of spatially localized molecules that have chemical groups that perform a specific function. Crucially, the resulting system is robust enough to be used for technological applications, such as in nano-sized components for sensor devices.

Currently, one of the best methods for creating patterned surfaces is optical lithography. In this process, a substrate is coated with a light-sensitive chemical (a photoresist), so that a design can be created by exposing the photoresist to ultraviolet light. The resulting pattern is used as a template for physical or chemical modification of the substrate. The design on the substrate is created by shining ultraviolet light through a 'mask', and herein lies the problem. The accuracy of the lithographic process is limited both by the wavelength of the light (which is usually relatively long — hundreds of nanometres) and the precision with which masks can be manufactured. The problem is partly alleviated by projecting the image from the mask through a system of lenses, to reduce the size of the image that reaches the photoresist. In this way, surfaces can be etched with designs at scales below 100 nanometres. But this still isn't small enough for proposed nanotechnology applications. Nevertheless, optical lithography is tremendously successful, and is used by the semiconductor industry to make microchips for computers.

For smaller surface features, only 'serial' methods are available, in which each part of the pattern is created one after the other, rather like drawing a design with a pen. These techniques include electron beam lithography (in which a pattern is formed in a resist using electron beams), or depositing molecules on a surface using the tip of an atomic force microscope (AFM). Patterns can be made with nanometre



**Figure 1 | Patterned arrays of molecules.** **a**, Thiol molecules self-assemble on gold surfaces to form monolayers of molecules. In this image, taken with a scanning tunnelling microscope, each dot represents a single thiol molecule (adamantanethiol). **b**, Madueno *et al.*<sup>1</sup> allowed a two-dimensional, porous network to self-assemble from two different kinds of aromatic molecules on a gold surface. On treatment of the network with adamantanethiol, monolayers form on the gold surface within the pores, yielding a regular pattern of hexagonal monolayers. Dark regions show the 'frame' of the molecular network. (Images courtesy M. Buck.)

resolution with these techniques, but they are impractical for applications involving large surface areas — although arrays of thousands of AFM tips have been used to deposit compounds across wider areas, in a process known as dip-pen lithography.

But there is another strategy. The smallest unit of a chemical pattern is an individual molecule, and appropriately designed molecules can self-assemble into molecular films with a regular order. An ideal approach would therefore be to use an ordered, self-assembled network as a template to arrange other molecules with nanometre-scale precision.

Compared with the self-assembly of macromolecular complexes in solution, the analogous process on a surface is less well developed<sup>2,3</sup>. Assemblies of molecules that form open network structures are of particular interest as surface templates because they contain cavities that can be filled by 'guest' molecules. A wide variety of molecules can in principle be accommodated, so that the overall properties of the networks could be modified simply by changing the guests. So far, only a few model systems have been studied, and in these systems the guests are simple molecules that aren't useful for any applications and don't greatly affect the properties of the network<sup>4-6</sup>. In any case, these model systems are too fragile to be used for nanotechnology applications.

Madueno *et al.*<sup>1</sup> now describe a robust, stable network of molecules that survives the formation of films (known as self-assembling monolayers, or SAMs) in its molecular corrals (Fig. 1). This provides a simple method for generating patterned SAMs of molecules that could, for example, bind specific biomolecules in a diagnostic application, or be used as a template for metal deposition. Furthermore, the authors show for the first time that such a network can be strong enough to survive further modification in solution. They also demonstrate that their network-SAM hybrid is stable in electrochemical experiments, which suggests that such structures could be processed to produce arrays of nanometre-sized structures made from metals or semiconductors.

But the implications of the work extend beyond practical applications — fundamental scientific questions could also be addressed using Madueno and colleagues' approach. If objects (molecules, atomic clusters or nanoparticles) can be attached to SAMs that are corralled in a network, this would allow the effects of confinement, at scales relevant to

quantum processes, to be studied. For example, how does such confinement affect SAM formation, and what is the smallest size at which they can form? The answers to these questions will be useful for determining how small certain nanotechnology devices could be.

One could also imagine narrowing down the size of the pores in a network until only a single object can be accommodated in each pore. The intermolecular distance between guest objects would then be controlled by the structure of the network. Such a system would be useful for studying the unfolding of polymers and proteins; energy transfers between molecules or nanoparticles; or confinement effects in chemical or enzymatic reactions. Another possibility would be to use the network–SAM hybrid as a template for three-dimensional self-assembled structures.

It remains to be seen how flexible self-assembling network–SAM structures will be. Madueno *et al.*<sup>1</sup> used gold substrates in their experiments, as this is the metal of choice for SAM formation. Alternative substrate materials have yet to be tried, but it is safe to assume that other metals will work. Many different SAM molecules are also available, each with different chemical groups attached. It is currently unclear whether all of these SAM molecules will be suitable as guests, however, as some of the appended groups might react with host molecules at random positions in the network, destroying the order of the resulting system. Finally, the molecules that form the network should also be investigated — to what extent can they be modified, or varied in size?

Madueno *et al.* have provided a basic recipe for patterning surfaces that allows the

promising possibilities of precisely delineated arrays of molecules to be explored. No doubt others will come up with their own variations with which to explore the limits of two-dimensional host–guest arrays, to eliminate imperfections and, perhaps most excitingly, to develop applications for them.

Michael Grunze is in the Department of Applied Physical Chemistry, University of Heidelberg, INF 253, 69120 Heidelberg, Germany.  
e-mail: michael.grunze@urz.uni-heidelberg.de

1. Madueno, R., Räisänen, M. T., Silien, C. & Buck, M. *Nature* **454**, 618–621 (2008).
2. Blunt, M. *et al. Chem. Commun.* 2304–2306 (2008).
3. Pawin, G., Wong, K. L., Kwon, K.-Y. & Bartels, L. *Science* **313**, 961–962 (2006).
4. Theobald, J. A. *et al. Nature* **424**, 1029–1031 (2003).
5. Stepanow, S., Lin, N., Barth, J. V. & Kern, K. *Chem. Commun.* 2153–2155 (2006).
6. Spillmann, H. *et al. Adv. Mater.* **18**, 275–279 (2006).

## CANCER

# Ins and outs of tumour control

Maria S. Soengas

**When a potentially dangerous cell can't be repaired, it must be either stopped or killed. Premature senescence of cancerous cells is one such 'stop' mechanism, in which immune mediators play an unexpected part.**

Various mutations can make normal mammalian cells cancerous. Luckily, several built-in mechanisms that inhibit tumour formation are in place. For example, a host of molecular factors, including tumour-suppressor proteins, blunt the accumulation of potentially dangerous cells by promoting either cellular senescence or apoptosis (programmed cell death)<sup>1</sup>. But tumour suppressors are neither infallible nor work in isolation. So the quest is on for other natural anticancer agents. In papers published in *Cell*, Kuilman *et al.*<sup>2</sup> and Acosta *et al.*<sup>3</sup> describe the role of several secreted immune mediators in promoting cellular senescence in response to oncogene activation.

Senescent cells are metabolically active and

can secrete various proteins, including growth factors, matrix metalloproteinases, protease inhibitors and cytokines, each of which can have multiple effects on the tumour micro-environment and, ultimately, on tumour development<sup>4,5</sup>. The two teams<sup>2,3</sup> also independently arrive at the secretome (a collection of secretory cellular molecules) as a central switch in oncogene-induced cellular senescence.

Kuilman *et al.* wanted to identify genes that respond differentially to the BRAF<sup>V600E</sup> oncogene, and came upon a footprint of immune mediators (cytokines and chemokines) that are induced only in senescent cells. For example, they find that, in fibroblasts expressing BRAF<sup>V600E</sup>, signals from this

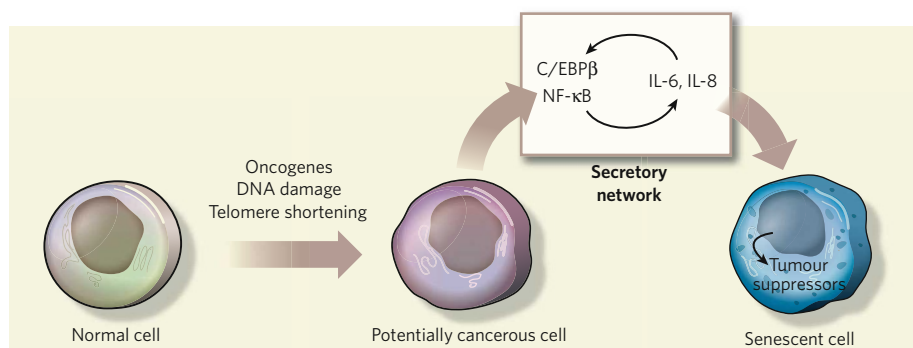
oncogene induce secretion of the cytokine IL-6, which is essential — although insufficient — for the induction and maintenance of senescence in the very cells that secrete them (Fig. 1).

The authors also find that the gene transcription factor C/EBP $\beta$  regulates IL-6 expression, and that it is crucial for mediating oncogene-induced senescence. Depletion of IL-6 reduced C/EBP $\beta$  expression and vice versa, suggesting that an interconnected feedforward loop operates between the genes encoding these proteins. The precise sequence of events downstream of IL-6 activity is undefined, but Kuilman and colleagues show that senescence driven by BRAF<sup>V600E</sup> depends, at least in part, on the induction of the tumour-suppressor protein p15<sup>INK4b</sup>. However, another tumour suppressor, p16<sup>INK4a</sup> — a senescence marker in many cell types — seems to be dispensable for BRAF<sup>V600E</sup> response in fibroblasts.

The cytokine IL-8 also probably responds to oncogenic signals to drive premature senescence<sup>2</sup>, specifically in benign tumours that express high levels of p16<sup>INK4a</sup>. Like IL-6, IL-8 can promote senescence in a p16<sup>INK4a</sup>-independent way in cultured cells<sup>2</sup>. It will be interesting to see whether that is the case *in vivo* — for example in the case of naevi (moles) driven by the BRAF<sup>V600E</sup> or RAS oncogenes, which may not necessarily induce the p16<sup>INK4a</sup> protein<sup>6</sup>.

Acosta *et al.*<sup>3</sup> also show the possible impact of the secretory network on oncogene-induced senescence. In search of factors that extend the lifespan of human fibroblasts, these authors performed a functional screen using a library of short-hairpin RNAs. They find that CXCR2 — a receptor for several CXC chemokines, including CXCL1–7 and IL-8 — is a central mediator of senescence. Depletion of CXCR2 delayed not only senescence driven by oncogenes and DNA damage, but also natural senescence due to the shortening of telomere sequences at the ends of chromosomes.

The sequence of events leading to the



**Figure 1 | Intracellular and secreted proteins control tumour development.** Oncogenic signals, DNA damage and telomere shortening can all turn normal cells into potentially cancerous ones. Two groups<sup>2,3</sup> find that such affected cells induce a secretory network (IL-6, IL-8, C/EBP $\beta$  and NF- $\kappa$ B), which in turn triggers expression of tumour-suppressor proteins, such as p15<sup>INK4b</sup>, p16<sup>INK4a</sup> and p53, and premature senescence to prevent tumorigenesis.



induction of senescence is as follows: oncogenes or attrition of telomeres induce signalling pathways that depend on the transcription factor NF- $\kappa$ B and the protein kinase p38. These two proteins coordinately induce the expression of CXCR2 and its ligands. CXCR2, in turn, favours accumulation of the tumour suppressor p53. Intriguingly, there is specificity in the inhibitory signals controlled by CXCR2, because other classic tumour suppressors (p21, p16<sup>INK4a</sup> and p19<sup>ARF</sup>) are not involved in CXCR2-controlled senescence.

Acosta *et al.*<sup>3</sup> further demonstrate high, albeit non-uniform, expression of CXCR2 in benign lesions in mice and humans. They identify a loss-of-function mutation in CXCR2 in a lung-cancer cell line. Although this mutation could be an artefact of cell culture, the observation serves as a proof-of-principle for the role of CXCR2 in tumour control.

It is noteworthy that IL-6, IL-8 and C/EBP $\beta$  have previously been linked to oncogene-driven senescence<sup>7–9</sup>. The new studies<sup>2,3</sup> provide mechanistic detail about the contribution of secreted proteins to cellular senescence, and show evidence for this activity *in vivo*. But secretory loops controlling premature senescence have other components. Another group<sup>10</sup> recently reported that the induction and secretion of insulin-like growth factor binding protein-7 (IGFBP7) is essential for BRAF<sup>V600E</sup>-induced senescence in melanocytes and fibroblasts. In fact, if expressed at high enough levels, IGFBP7 is sufficient to drive premature senescence.

In contrast to classic tumour suppressors, which are inactivated or downregulated during tumorigenesis, the expression of IL-6, IL-8 or CXCR2 is not lost in advanced tumours. Therefore, it seems likely that the wiring of the secretome is fundamentally different in normal and tumour cells. In senescent cells, oncogenes induce some secreted proteins at levels 30–100 times higher than normal. This substantial increase in expression may saturate the protein-folding capacity of the endoplasmic reticulum, and the associated stress that this causes — the unfolded protein response — can also contribute to senescence<sup>11</sup>.

These findings' significance is directly related to their clinical implications. Together, the three studies<sup>2,3,10</sup> describe more than 20 proteins whose inactivation bypasses oncogene-induced senescence. The immediate corollary is that there are at least 20 previously unknown genes that cancer cells can target to evade both natural suppressive mechanisms and the potential beneficial effect of treatments that depend on the activation of tumour-cell senescence.

What's more, Kuilman *et al.*<sup>2</sup> describe changes within discrete nodes in the cytokine network — for example, deregulation of IL-6 or C/EBP $\beta$  — that might be sufficient to revert senescence to a proliferative state, without the need for other genetic alterations. Nonetheless, it may still be possible to modulate the balance in the cytokine network in order to

induce tumour-cell death by apoptosis. This pro-apoptotic activity of secreted proteins was demonstrated by forced expression of IGFBP7 in melanoma cells<sup>10</sup>. Understanding how to interfere with the cytokine secretome to efficiently kill advanced tumours, without activating compensatory mechanisms, is a challenge that will undoubtedly lead us to other unexpected loops and kinks in cancer cells and their microenvironment.

Maria S. Soengas is in the Melanoma Group, Molecular Pathology Programme, Spanish National Cancer Research Centre, Madrid 28029, Spain. e-mail: msoengas@cniio.es

1. Finkel, T., Serrano, M. & Blasco, M. A. *Nature* **448**, 767–774 (2007).
2. Kuilman, T. *et al.* *Cell* **133**, 1019–1031 (2008).
3. Acosta, J. C. *et al.* *Cell* **133**, 1006–1018 (2008).
4. Campisi, J. & d'Adda di Fagagna, F. *Nature Rev. Mol. Cell Biol.* **8**, 729–740 (2007).
5. Chien, Y. & Lowe, S. W. *Cell* **132**, 339–341 (2008).
6. Michaloglou, C. *et al.* *Nature* **436**, 720–724 (2005).
7. Mason, D. X., Jackson, T. J. & Lin, A. W. *Oncogene* **23**, 9238–9246 (2004).
8. Schnabl, B., Purbeck, C. A., Choi, Y. H., Hagedorn, C. H. & Brenner, D. *Hepatology* **37**, 653–664 (2003).
9. Sebastian, T., Malik, R., Thomas, S., Sage, J. & Johnson, P. F. *EMBO J.* **24**, 3301–3312 (2005).
10. Wajapeyee, N., Serra, R. W., Zhu, X., Mahalingam, M. & Green, M. R. *Cell* **132**, 363–374 (2008).
11. Denoyelle, C. *et al.* *Nature Cell Biol.* **8**, 1053–1063 (2006).

## PLANETARY SCIENCE

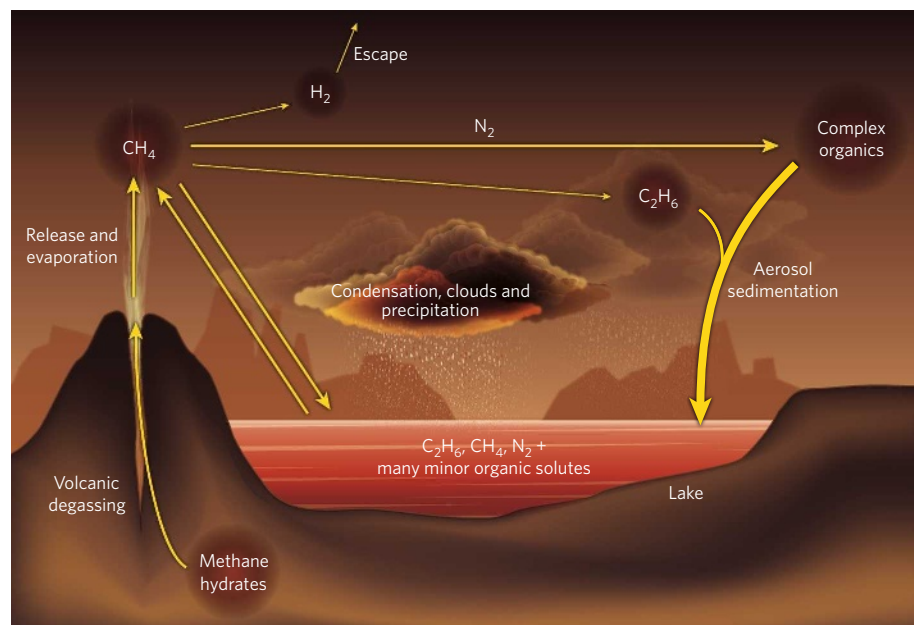
# Organic lakes on Titan

François Raulin

**While orbiting Saturn, the Cassini spacecraft has spotted lakes containing ethane on Titan, the planet's largest moon. Titan is so far the only planetary object other than Earth that is known to have liquid bodies on its surface.**

Titan is the largest satellite of the giant planet Saturn. Its diameter of 5,150 kilometres makes it bigger than Mercury and only 25% smaller than Mars. It is the second-largest satellite in the Solar System after Jupiter's Ganymede based on diameter alone, and would claim first place if its dense atmosphere, which extends more than 1,000 kilometres above its surface, were included. Fly-bys of Titan by the Voyager spacecraft in the early 1980s helped

determine the main composition, temperature and pressure of this atmosphere. Although much colder than Earth's (around 90–94 kelvin at the surface), and lacking molecular oxygen (O<sub>2</sub>), it shows many similarities to the atmosphere of our planet. It consists mainly of molecular nitrogen (N<sub>2</sub>), with a surface pressure of 1.5 bar and, like Earth, it has a structure comprising a troposphere and a stratosphere. On page 607 of this issue, Brown *et al.*<sup>1</sup> confirm



**Figure 1 | Titan's methane/ethane cycle.** Methane (CH<sub>4</sub>) is released into the atmosphere from Titan's interior stores through volcanic action, and evaporates from the lakes of methane and ethane (C<sub>2</sub>H<sub>6</sub>) identified by the Cassini spacecraft on the satellite's surface<sup>1</sup>. Chemical reactions in the atmosphere convert it to ethane; complex organic aerosols consisting of carbon, hydrogen and nitrogen; and hydrogen gas (H<sub>2</sub>), which escapes into space. Ethane and methane partly condense, forming clouds and hazes that precipitate, replenishing the lakes and bearing many organic species in solution.

the existence of a further similarity between Earth and Titan — the presence of open bodies of liquid on its surface.

The idea that liquid bodies exist on Titan's surface is not new. After the discovery that Titan's atmosphere contains a substantial amount of methane, the question of its origin and evolution was seen as a key to understanding Titan's mysteries. A combination of light from the Sun and high-energy electrons from Saturn's magnetosphere would have destroyed all methane in the atmosphere within a few tens of millions of years, implying that the methane was being replenished by a source on, or in, Titan. The main volatile organic product of the light-induced breakdown of methane is ethane, and it was proposed that an ocean consisting of liquid ethane and methane with dissolved N<sub>2</sub> covered Titan<sup>2</sup>.

The Voyager probes were unable to get a peek at Titan's surface, as it was masked by hazy layers in the atmosphere. But with the Cassini-Huygens mission, this veil has finally been lifted. Several instruments on the Cassini spacecraft have been observing the moon in regions of the electromagnetic spectrum in which the atmosphere is transparent.

These observations quickly ruled out the possibility of a global ocean<sup>3</sup>, but radar data strongly supported the possibility of smaller seas and lakes, mainly in the colder northern regions<sup>4</sup>. Furthermore, camera images taken in mid-2005 showed a dark surface feature near the south pole as big as North America's Lake Ontario, speculatively named Ontario Lacus. Brown *et al.*<sup>1</sup> have now identified the characteristic spectrum of liquid ethane in infrared spectra of Ontario Lacus.

The Cassini-Huygens mission has demonstrated the existence of a complex cycle of methane in Titan's geochemistry similar to the water cycle on Earth (Fig. 1). Large stores of methane seem to exist within Titan in the form of methane hydrates (clathrates) trapped during the formation of the satellite<sup>5</sup>. Methane may also be produced through the reaction of water with igneous rocks under high pressures, which would also form hydrogen gas. Ethane, formed by photodissociation of methane in the atmosphere, accumulates on Titan's surface, replenishing the surface lakes — which are therefore one possible ethane reservoir, as are haze particles in the atmosphere into which ethane can be sequestered<sup>6</sup>.

Titan's lakes are probably a liquid ethane-methane mixture together with dissolved nitrogen, as previously proposed for the speculative oceans<sup>2</sup>. Also dissolved in them will be a variety of solutes, mainly organic compounds produced in the atmosphere and rained down in aerosol particles consisting of a nucleus of macromolecular materials coated with volatile hydrocarbons and nitriles<sup>7</sup>. Such solutes will be much more concentrated in the lakes than in the atmosphere<sup>8</sup>. Despite the low temperatures, the action of high-energy cosmic rays reaching the satellite's surface may produce additional organic

compounds in this exotic chemical reactor.

It has been suggested that cold liquids such as those found in the lakes of Titan could contain life<sup>9</sup>, but they also have less speculative interests for astrobiologists<sup>10</sup>. Although composed of a low-temperature, nonpolar solvent very different from water, they provide an analogue to Earth's oceans and are a potential chemical reactor. It is to be hoped that these organic lakes are made a priority target for future exploration missions, such as the Titan/Saturn System Mission (TSSM), a joint venture between NASA and the European Space Agency currently undergoing feasibility studies. ■

François Raulin is at the Université Paris 12 Val de Marne, 61 avenue du Général de Gaulle, 94010 Créteil Cedex, France.  
e-mail: raulin@lisa.univ-paris12.fr

1. Brown, R. H. *et al.* *Nature* **454**, 607–610 (2008).
2. Lunine, J. I., Stevenson, D. J. & Yung, Y. L. *Science* **222**, 1229–1230 (1983).
3. Porco, C. C. *et al.* *Nature* **434**, 159–168 (2005).
4. Stofan, E. R. *et al.* *Nature* **445**, 61–64 (2007).
5. Niemann, H. B. *et al.* *Nature* **438**, 779–784 (2005).
6. Hunten, D. M. *Nature* **443**, 669–670 (2006).
7. Israël, G. I. *et al.* *Nature* **438**, 796–799 (2005).
8. Dubouloz, N. *et al.* *Icarus* **82**, 81–96 (1989).
9. McKay, C. P. & Smith, H. D. *Icarus* **178**, 274–276 (2005).
10. Raulin, F. *Space Science Rev.* doi:10.1007/s11214-006-9133-7 (2007).

## BEHAVIOURAL NEUROSCIENCE

# The circuit of fear

Pankaj Sah and R. Frederick Westbrook

**Do you find it difficult to overcome an irrational fear? Blame it on the specific neural circuits hardwired in the brain that control fear recognition, and fear renewal even when fear has seemingly been overcome.**

Learning to predict danger allows animals to defend themselves against harm and is crucial for survival. The neural mechanisms that subserve these functions are evolutionarily old, and their dysfunction is thought to underlie a host of anxiety disorders in humans, including post-traumatic stress and panic disorder<sup>1</sup>. Two papers in this issue, by Herry *et al.*<sup>2</sup> (page 600) and Likhtik *et al.*<sup>3</sup> (page 642), pinpoint the neural circuits that mediate the bidirectional transition between a defensive behaviour — fear — and the default exploratory behaviour.

To study the neural mechanisms that mediate fear responses, fear conditioning is widely used. In a typical procedure, animals are exposed to a normally harmless stimulus (such as a sound or light) before a brief exposure to an aversive stimulus — typically a foot shock. A few such rounds of pairing the harmless (conditioned) stimulus with the aversive (unconditioned) stimulus create an association between them in the animals' minds. So, on subsequent exposure to only the conditioned stimulus, they remember the association and express behavioural and physiological responses that correspond to fear in humans.

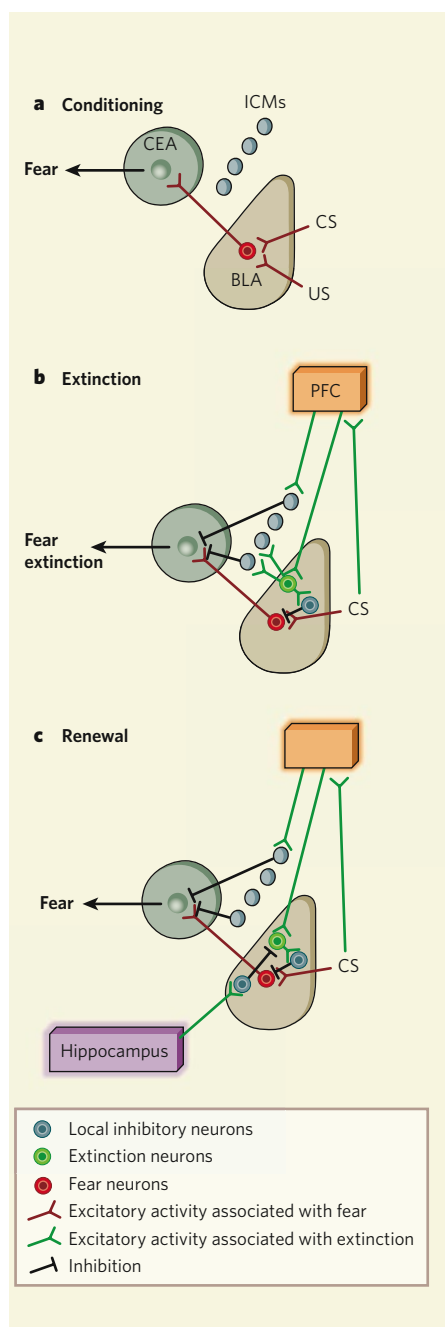
It is important to study how fear is first learned. However, the pertinent question for clinicians is how fear can be eliminated or reduced. Extinction of fear occurs when the association between the conditioned and the unconditioned stimuli is broken by repeated presentations of the conditioned stimulus only<sup>4</sup>. But does the association get completely erased from the memory? The answer is no. Although the conditioned stimulus eventually fails to elicit fear responses, much, if not all, of the original learned fear survives extinction<sup>5</sup>. So when the extinguished conditioned stimulus

is tested either during, or shortly after, exposure to a dangerous context again, the conditioned fear is renewed spontaneously. Extinction therefore involves new learning, and its activation by situational cues inhibits the expression of fear responses to the conditioned stimulus.

Three brain regions implicated in fear conditioning and extinction are the amygdala, the medial prefrontal cortex and the hippocampus. The amygdala is the central structure involved in both fear conditioning and extinction<sup>1</sup>. It is divided into two main compartments: the basolateral amygdala, which receives sensory information about the conditioned and the unconditioned stimuli; and the central nucleus, which receives information processed in the basolateral amygdala. Neuronal projections from the central nucleus to the hypothalamus and brainstem then evoke the behavioural and physiological responses that constitute fear<sup>6</sup> (Fig. 1a, overleaf).

The learning of fear involves the co-occurrence of weak (conditioned stimulus) and strong (unconditioned) inputs to single neurons and leads to enhancement of neuronal activity known as excitatory post-synaptic potentials, which are elicited by the conditioned stimulus and mediated by NMDA glutamate receptors. This synaptic plasticity results in the conditioned stimulus evoking an increased response in basolateral-amygdala neurons that project to fear-output circuits in the central nucleus. Extinction results from NMDA-receptor-mediated changes in the amygdala<sup>7</sup> that block the effects of this long-term potentiation, possibly by activating local neurons that secrete the inhibitory neurotransmitter GABA. The medial prefrontal cortex mediates consolidation of extinction,





**Figure 1 | Neural circuits that mediate fear.** **a**, During fear conditioning, convergence of inputs from the conditioned stimulus (CS) and unconditioned stimulus (US) to fear neurons in the basolateral amygdala (BLA) leads to potentiation of the conditioned input and activation by the CS of neurons in the central nucleus (CEA) that initiates physiological and behavioural responses characteristic of fear. **b**, During extinction, inputs from the medial prefrontal cortex (PFC) activate neurons in the intercalated cell masses (ICMs) — either directly or through activation of extinction neurons in the basolateral amygdala — which then inhibit the activity of fear output neurons in the CEA. **c**, During fear renewal, inputs from the hippocampus, which evaluates the current context, activate inhibitory interneurons in the basolateral amygdala that silence extinction neurons, thus restoring fear responses.

as well as the subsequent activation of these GABA-secreting neurons by the conditioned stimulus. The hippocampus probably regulates the situational control over this activation, and so over extinction. But how fear-conditioning and extinction memories are stored, and how they are translated into behavioural outcomes, is poorly understood.

Herry *et al.*<sup>2</sup> set out to examine the role of the basolateral amygdala in fear learning and extinction in mice. In agreement with previous work<sup>8,9</sup>, they found that the pairing of conditioned and unconditioned stimuli results in enhanced firing of a population of basolateral-amygdala neurons — fear neurons — when animals are subsequently exposed to the conditioned stimulus alone. But, intriguingly, another population of neurons — extinction neurons — emerged, which selectively respond to the conditioned stimulus undergoing extinction. Moreover, moving the mice from the extinction context to the conditioning context not only renews fear responses but also leads to reactivation of fear neurons and deactivation of the extinction neurons.

The authors show that the change from fear-neuron firing to extinction-neuron firing precedes the behavioural shift from a fear to a no-fear response, suggesting that activities of these neurons control the observed behaviour. In addition, infusion of the GABA-receptor agonist muscimol into the basolateral amygdala to inactivate neurons there impaired both extinction learning and fear renewal, indicating that these neurons' activity is crucial for the inhibition that underlies extinction.

To examine connectivity of fear neurons and extinction neurons, Herry *et al.* next stimulated the medial prefrontal cortex and the hippocampus. Their findings suggest that fear neurons receive inputs from the hippocampus and project to the medial prefrontal cortex. Extinction neurons, by contrast, are connected to the medial prefrontal cortex in both directions.

How does the extinction neurons' activity produce the shift from fear to safety? Between the basolateral amygdala and the central nucleus are clusters of inhibitory neurons known as the intercalated cell masses (ICMs). These neurons are innervated by input neurons from the basolateral amygdala, and their activity inhibits output neurons in the central amygdala by shunting dendritic excitatory inputs<sup>10</sup>. Neurons in the medial prefrontal cortex also densely innervate ICM neurons<sup>11</sup>. It has been suggested<sup>12,13</sup>, therefore, that activation of ICM neurons by these cortical neurons inhibits the amygdala's output during extinction, but conclusive evidence has been lacking.

Likhtik *et al.*<sup>3</sup> used an innovative technique to selectively silence ICM neurons. Following fear conditioning and its extinction in rats, the authors delivered a toxin called saporin specifically to the animals' ICM neurons, to block protein synthesis and cause ICM cell

death. They find that, with a decrease in ICM neurons, extinguished fear responses can once again be reactivated, suggesting that these neurons are required for the expression of learned extinction (Fig. 1).

These authors' work, together with that of Herry *et al.*<sup>2</sup>, clearly defines some of the neural circuits that mediate the extinction of conditioned fear. The amygdala, together with the hippocampus and the medial prefrontal cortex, uses situational information to evaluate an extinguished fear-conditioned stimulus and acts as a switching circuit that guides the appropriate behavioural response (Fig. 1). But, as with all interesting studies, these observations also raise questions.

The existence of fear and extinction neurons in the basolateral amygdala has been inferred from changes in the activity in this region after learning fear and extinction. These neurons are yet to be identified. Do extinction neurons activate ICM cells directly or through projections to the medial prefrontal cortex? Inhibitory neurons in the basolateral amygdala strongly inhibit the activity of local pyramidal neurons<sup>14</sup>.

Does inhibition of fear neurons during extinction result from activation of such local feedback interneurons in the basolateral amygdala? Herry and colleagues' observations suggest that long-term memories of both learned fear and extinction are not ultimately stored in the amygdala. So another question is how these learned behaviours are 'gated' after consolidation.

Answers to these questions will require exploration of the local circuitry of the amygdala, identifying fear and extinction neurons and uncovering the final targets of these cells. Neuronal receptors in these circuits — such as those targeted with saporin in Likhtik and colleagues' study — are likely to become targets for the development of specific treatments for many anxiety disorders.

Pankaj Sah is at the Queensland Brain Institute, University of Queensland, St Lucia, Queensland 4072, Australia. R. Frederick Westbrook is in the School of Psychology, University of New South Wales, Sydney, New South Wales 2052, Australia.

e-mail: pankaj.sah@uq.edu.au

1. Davis, M. *Annu. Rev. Neurosci.* **15**, 353–375 (1992).
2. Herry, C. *et al. Nature* **454**, 600–606 (2008).
3. Likhtik, E., Popa, D., Apergis-Schoute, J., Fidacaro, G. A. Jr & Paré, D. *Nature* **454**, 642–645 (2008).
4. Pavlov, I. P. *Conditioned Reflexes* (Dover, 1927).
5. Rescorla, R. A. *Annu. Rev. Neurosci.* **11**, 329–352 (1988).
6. Sah, P., Faber, E. S. L., Lopez de Armentia, M. & Power, J. *Physiol. Rev.* **83**, 803–834 (2003).
7. Falls, W. A., Miserendino, M. J. & Davis, M. J. *Neurosci.* **12**, 854–863 (1992).
8. Quirk, G. J., Repa, J. C. & LeDoux, J. E. *Neuron* **15**, 1029–1039 (1995).
9. Collins, D. R. & Paré, D. *Learn. Mem.* **7**, 97–103 (2000).
10. Delaney, A. J. & Sah, P. J. *Neurophysiol.* **86**, 717–723 (2001).
11. McDonald, A. J., Mascagni, F. & Guo, L. *Neuroscience* **71**, 55–75 (1996).
12. Paré, D., Quirk, G. J. & LeDoux, J. E. *J. Neurophysiol.* **92**, 1–9 (2004).
13. Quirk, G. J., Likhtik, E., Pelletier, J. G. & Paré, D. *J. Neurosci.* **23**, 8800–8807 (2003).
14. Woodruff, A. R. & Sah, P. J. *Neurophysiol.* **98**, 2956–2961 (2007).

## MATERIALS SCIENCE

# A desirable wind up

Neil Mathur

**Multiferroic materials have complex domain structures that make precise interpretation of their behaviours difficult. The creation and study of a single-domain multiferroic crystal is thus an exciting development.**

The terms multiferroic and magnetoelectric are often confused. Materials are multiferroic if they have both spontaneously ordered magnetic (ferromagnets or antiferromagnets) and electrical dipole moments (ferroelectrics). However, spontaneous order is not required for the two dipole species to exhibit magnetoelectric coupling, a phenomenon that enables materials to interconvert magnetic and electrical signals. Most multiferroic materials are expected to display some form of magnetoelectric coupling, but the formation of multiple magnetic and electrical domains complicates the analysis of magnetoelectric effects. Writing in *Physical Review Letters*, Lebeugle *et al.*<sup>1</sup> report on the multiferroic and magnetoelectric properties of single crystals of bismuth ferrite ( $\text{BiFeO}_3$ ) that contain just one electrical and magnetic domain — possibly because the crystals were made at low enough temperatures to be ferroelectric once formed.

Multiferroics that are both ferromagnetic and ferroelectric are unfortunately rather rare. Ferroelectric antiferromagnets are much more common and are also described as multiferroic. These ferroelectrics have no overall magnetization because the magnetic dipoles in adjacent unit cells cancel out.  $\text{BiFeO}_3$  is unusual because it is a ferroelectric antiferromagnet at relatively high temperatures, up to and above room temperature<sup>1</sup>. It sprang to prominence in 2003 following claims<sup>2</sup> that the strain present in thin films enhances its ferroelectric polarization and generates significant magnetization. Both claims were incorrect. In the bulk, a large electrical polarization had previously been expected<sup>3</sup> and was subsequently observed<sup>4</sup>; and the large magnetization in films could not be reproduced<sup>5</sup> despite the possibility of contributions from ferromagnetic domain walls<sup>6</sup>.

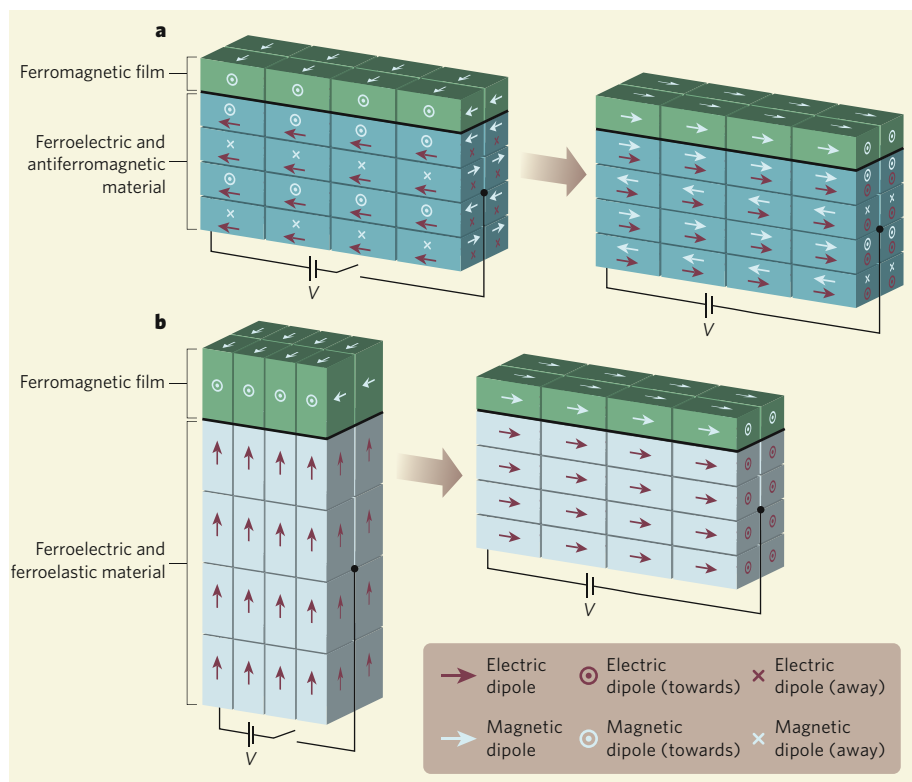
In the long term,  $\text{BiFeO}_3$  may still be technologically interesting for two reasons. First, it represents a lead-free alternative to lead zirconium titanate — currently the material of choice for ferroelectric random access memory (FeRAM) — although there are serious concerns about electrically induced breakdown<sup>7</sup>. Second,  $\text{BiFeO}_3$  could be used for voltage-driven (low-power) magnetic write operations, for example in magnetic random access memory (MRAM). In this scenario (Fig. 1a), an electric field would magnetoelectrically switch the antiferromagnetic order of a layer of  $\text{BiFeO}_3$ , which would, in turn, switch the magnetization of an overlying thin ferromagnetic film by exploiting a process known as exchange-

coupling.  $\text{BiFeO}_3$  is already known to fulfil two of the requirements for such switching: its complex antiferromagnetic order<sup>1</sup> can modify the magnetically driven magnetic switching of an adjacent ferromagnetic film due to exchange bias<sup>8</sup>; and this antiferromagnetic order can be electrically switched in thin films<sup>9</sup>.

Lebeugle *et al.*<sup>1</sup> have shown that electrically induced magnetic switching also occurs in bulk  $\text{BiFeO}_3$ , by using neutron scattering to observe the electrically induced conversion of a single antiferromagnetic domain into a multi-domain state. This is a remarkable discovery because strong magnetoelectric coupling had been thought impossible in  $\text{BiFeO}_3$  owing to the superimposition of a cycloidal variation

onto the antiferromagnetically ordered dipoles. Although Lebeugle *et al.* present the first unambiguous demonstration that the long-period variation may indeed be a cycloid (with a period of 64 nanometres), they argue that the strong magnetoelectric coupling arises locally even though it is globally forbidden by this cycloid. It was previously thought that strong magnetoelectric coupling would be possible<sup>9</sup> in  $\text{BiFeO}_3$  if the cycloid were 'unwound' using the strain present in thin films<sup>10</sup>. But Lebeugle *et al.* suggest that unwinding orients the magnetic and electric dipoles in a pattern that is in fact less favourable for coupling and so should be avoided.

In a recent study<sup>11</sup>, the magnetic domains in an exchange-biased ferromagnetic film of cobalt doped with iron ( $\text{Co}_{0.9}\text{Fe}_{0.1}$ ) were controlled by the application of a lateral electric field to an underlying layer of  $\text{BiFeO}_3$ . The switching of these magnetic domains may have been due to the desired combination of magnetoelectric coupling in the  $\text{BiFeO}_3$  and exchange coupling between the two materials (Fig. 1a). However,  $\text{BiFeO}_3$  is ferroelastic, and therefore the strain associated with ferroelectric switching could be wholly or partially



**Figure 1 | Electrical control of ferromagnetic films.** **a**, An electrically driven change in ferroelectric polarization and thus antiferromagnetic order, as demonstrated in a more complex geometry for  $\text{BiFeO}_3$  by Lebeugle *et al.*<sup>1</sup>, could switch the magnetization of an overlying thin ferromagnetic film by quantum-mechanical exchange. **b**, An electrically driven change in the electrical polarization of a ferroelectric that is also ferroelastic produces mechanical strain, which is transmitted to an overlying ferromagnetic thin film. This deformation modifies the preferred orientations of the magnetic dipoles and therefore the macroscopic magnetization<sup>12,13</sup>. For ferroelectrics with both antiferromagnetic and ferroelastic properties, such as  $\text{BiFeO}_3$ , it is difficult to distinguish between these two mechanisms. Only a few unit cells within a single ferroic domain near the interface are shown, the unit cell aspect ratios are greatly exaggerated and the electrodes are depicted as points. The possibility of short-circuits through the films, which are likely to be metallic, can be avoided if the electric field is arranged perpendicular<sup>12,13</sup>, rather than parallel<sup>11</sup>, to the interface.



responsible (Fig. 1b). Strain is an established means by which to achieve magnetoelectric coupling between two suitable materials that are intimately connected<sup>12,13</sup>.

Even in an experiment in which it was clear that exchange bias (from yttrium manganite,  $\text{YMnO}_3$ , rather than  $\text{BiFeO}_3$ ) could be electrically controlled<sup>14</sup>, it was not clear whether the observed suppression in exchange bias was due to magnetoelectric coupling, strain, heating or a combination of all three. Experiments without strain and heating would permit the mechanism shown in Figure 1a to be unambiguously identified. Strain and heating are also potentially detrimental to device performance. Strain could be abolished either by clamping active regions by their surrounds, or by using a ferroelectric that is not ferroelastic and therefore does not change shape when electrically switched. Heating could be measured by monitoring leakage currents through the material.

The control of magnetic domains in a ferromagnetic film through an adjacent voltage-driven multiferroic promises to combine the best features of FeRAM and MRAM. But this promise is very distant, with many challenges beyond the unequivocal identification of the mechanism at buried and imperfect interfaces. One initial challenge would be to vary the crystallographic orientation of the  $\text{BiFeO}_3$  with respect to the applied electric field.

The clear link established by Lebeugle *et al.*<sup>1</sup> between the structural, electronic and magnetic properties of a complex oxide is refreshing. Furthermore, their approach of starting with a single bulk domain relied strongly on the frequently under-appreciated activity of sample preparation. Unexpectedly, the cycloid seems to be good for magnetoelectric coupling, and  $\text{BiFeO}_3$  may yet provide further surprises given its many phase transitions at both high<sup>15</sup> and low<sup>16</sup> temperatures. ■

Neil Mathur is in the Department of Materials Science, New Museums Site, Pembroke Street, Cambridge CB2 3QZ, UK.  
e-mail: ndm12@cam.ac.uk

## DEVELOPMENTAL BIOLOGY

# Our fly cousins' gut

Chrysoula Pitsouli and Norbert Perrimon

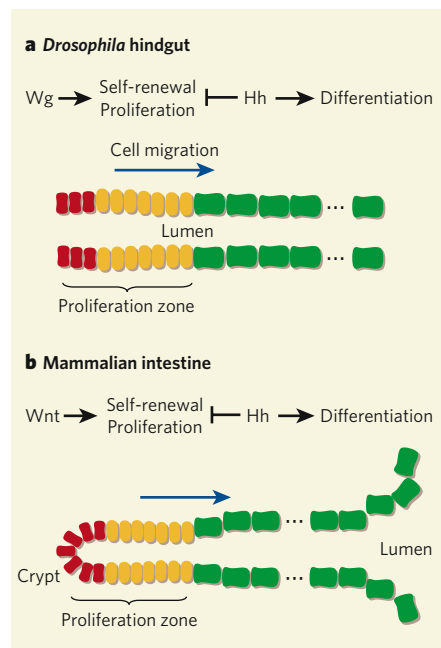
**What do we humans have in common with flies? Quite a lot, at least at the cellular and molecular levels. Our intestine, for instance, is similar to that of the fly, not only in function but also in its development and maintenance.**

Epithelial cells lining the intestine become damaged by ingested food, pathogens and toxins, and so must be constantly renewed. Intestinal stem cells compensate for the loss of these cells by producing all of the mature intestinal epithelial cell types<sup>1</sup>, and an imbalance in this process can lead to diseases such as cancer<sup>2</sup>. Our knowledge about intestinal homeostasis comes mainly from studies in mice. But the fact that flies are so amenable to genetic dissection — a task more difficult to carry out in mammals — has also allowed fascinating insights<sup>3–5</sup> into this process. The latest offering comes from a study by Takashima *et al.*<sup>6</sup> (page 651 in this issue), who highlight a remarkable similarity between mammals and the fruitfly *Drosophila melanogaster* in intestinal stem-cell renewal and differentiation, and the underlying molecular signalling pathways.

In mice and humans, intestinal stem cells are located in the numerous invaginations of the gut epithelium called crypts. The mesenchyme tissue surrounding the crypts acts as a niche for these cells, and guides their decision to self-renew or differentiate. At the bottom of each crypt, stem cells self-renew and produce progeny known as transit-amplifying cells, which divide several times before differentiating into the various mature intestinal cell types. During this process, the cells are continually migrating from the bottom of the crypts towards the gut lumen. There, the mature intestinal cells are exposed to the gut content and are finally shed from the epithelium<sup>1</sup>.

The digestive system of adult *Drosophila* is a tube composed of epithelial cells with absorptive and secretory functions similar to those of the mammalian intestine<sup>7</sup>. A population of stem cells interspersed among the epithelial cells in the insect's midgut maintains homeostasis there<sup>3–5</sup>, dividing to replenish epithelial cells lost through apoptosis (programmed cell death)<sup>4</sup>. Takashima *et al.*<sup>6</sup> now identify a different population of cells that they propose represents stem cells responsible for the development and maintenance of the insect's hindgut.

The authors show that the way in which homeostasis is achieved in the hindgut of adult *Drosophila* is strikingly similar to the related process in a mammalian crypt. By studying cell morphology in the hindgut epithelium and monitoring the kinetics of cell proliferation, they find that the anterior part of the hindgut (that nearest to the midgut) corresponds to a proliferation zone. This zone consists of anteriorly located, slowly proliferating cells, which



**Figure 1 | Similarities between the fly hindgut and a mammalian intestinal crypt.** Takashima *et al.*<sup>6</sup> propose that, in the *Drosophila* hindgut (a), stem cells (red), proliferating progenitor cells (yellow) and differentiated intestinal cells (green) are arrayed similarly to those in a crypt of the mammalian intestine (b). In both settings, the Wnt/Wg and Hedgehog (Hh) signals seem to be responsible for self-renewal/proliferation and differentiation, respectively. Self-renewal and differentiation of cells in adult *Drosophila* hindgut seem to be reminiscent of a process that occurs during the larval development of this organism.

give rise to fast-proliferating progeny (reminiscent of transit-amplifying cells in mammals) towards the posterior. These cells in turn produce differentiated hindgut cells, which lie even more posteriorly. Lineage analysis of the adult hindgut led the authors to propose that the slowly proliferating cells behave as stem cells (they self-renew and produce differentiated progeny).

Furthermore, this proliferation zone seems to have a crucial role in the development of the hindgut. Takashima *et al.*<sup>6</sup> traced hindgut cells from their development in larvae, through metamorphosis to adult flies. They find that, in the larva, these cells proliferate to generate adult tissue by replacing the dying larval cells. But they differ from other types of adult fly progenitor cell in that they persist in the adult insect, proliferating and constitutively

1. Lebeugle, D. *et al.* *Phys. Rev. Lett.* **100**, 227602 (2008).
2. Wang, J. *et al.* *Science* **299**, 1719–1722 (2003).
3. Teague, J. R., Gerson, R. & James, W. J. *Solid State Commun.* **8**, 1073–1074 (1970).
4. Lebeugle, D., Colson, D., Forget, A. & Viret, M. *Appl. Phys. Lett.* **91**, 022907 (2007).
5. Eerenstein, W. *et al.* *Science* **307**, 1203a (2005).
6. Daraktchiev, M., Catalan, G. & Scott, J. F. *Ferroelectrics* (in the press).
7. Lou, X. J. *et al.* *Appl. Phys. Lett.* **90**, 262908 (2007).
8. Dho, J., Qi, X., Kim, H., MacManus-Driscoll, J. L. & Blamire, M. G. *Adv. Mater.* **18**, 1445–1448 (2006).
9. Zhao, T. *et al.* *Nature Mater.* **5**, 823–829 (2006).
10. Béa, H., Bibes, M., Petit, S., Kreisel, J. & Barthélémy, A. *Phil. Mag. Lett.* **87**, 165–174 (2007).
11. Chu, Y.-H. *et al.* *Nature Mater.* **7**, 478–482 (2008).
12. Thiele, C., Dörr, K., Bilani, O., Rödel, J. & Schultz, L. *Phys. Rev. B* **75**, 054408 (2007).
13. Eerenstein, W., Wiora, M., Prieto, J. L., Scott, J. F. & Mathur, N. D. *Nature Mater.* **6**, 348–351 (2007).
14. Laukhin, V. *et al.* *Phys. Rev. Lett.* **97**, 227201 (2006).
15. Polomska, M., Kaczmarek, K. & Paják, Z. *Phys. Stat. Sol. (a)* **23**, 567–574 (1974).
16. Singh, M. K., Katiyar, R. S. & Scott, J. F. *J. Phys. Condens. Matter* **20**, 252203 (2008).

## BIOMOLECULAR ENGINEERING

## Negative success in tiny tree

When engineers were building beam engines in the early eighteenth century to pump out water-logged mines, they found that they couldn't pull water up more than about 9 metres (the height of water that can be supported by the drop in pressure between the atmosphere and a vacuum). Trees grow many times taller — more than 100 metres in the case of the tallest redwoods. Yet they supply their leaves with a constant flow of water. They achieve this feat by keeping the water high up in their trunks under pressures many atmospheres below that of a vacuum.

Elsewhere in this issue, Wheeler and Stroock report a duplication of this trick: they have created a tiny 'synthetic tree' through whose trunk water flows at pressures of around -10 atmospheres (T. D. Wheeler and A. D. Stroock *Nature* **455**, 208–212; 2008).

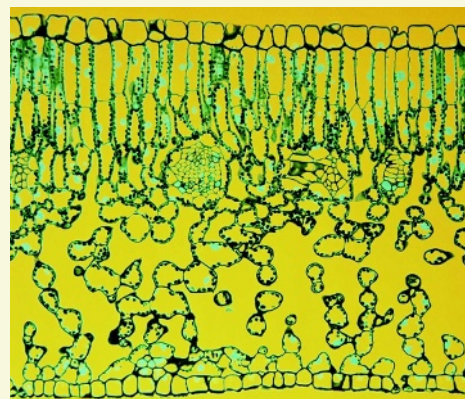
In trees, evaporation of water from leaf cells called spongy mesophyll pulls water up through hollow cells

in the trunk (spongy mesophyll is the tissue in the lower half of this picture, a cross-section through a leaf). The strong, cohesive properties of water, responsible for its powerful surface tension, allow the water to exist at large negative pressures. But even the smallest bubble would explosively expand into the water, disrupting its flow in a process known as cavitation. The interface between the plant's water system and the air, formed by the spongy mesophyll, must allow water to pass, but not the gas molecules that would cause cavitation.

To create their tree, Wheeler and Stroock use a hydrogel, which mimics the mesophyll by holding water in molecular-scale pores, smaller than those of other porous solids. As their respective 'root' and 'leaf', the authors formed two networks of channels, 10 micrometres in diameter, in a sheet of poly(hydroxyethyl methacrylate), and connected them

by a single channel, the 'trunk'. With the 'root' exposed to a source of water and the 'leaf' to a stream of damp air, water flows through the system powered solely by 'leaf' evaporation. The pressures developed in the trunk are some 15 times more negative than in any previously reported pumping system.

The device is shown in Figure 3a of the paper (page 210). It is just 5 centimetres long, and the flow is a little over 2 micrograms of water per second — but from such small acorns do mighty oaks grow. The synthetic tree can provide a test device for theories of tree physiology and, scaled-up, the technology could find uses in passive pumps or cooling devices — evaporation



makes the 'leaf' a heat sink. Also, the large negative pressures developed might be used to drag water out of even quite dry soils, simultaneously filtering out impurities by passage through the 'root' hydrogel. This process, which the authors dub "reverse reverse osmosis", could form the basis of solar-powered mining of pure water in arid or contaminated environments.

**Christopher Surridge**

largest such entangled states contain eight ions<sup>7</sup> or six photons<sup>8</sup>. By contrast, experiments with entangled states that involve a large particle number (greater than 100) lack the possibility of controlling individual particles, which is necessary to exploit multipartite correlations. For example, nearest-neighbour interactions of atoms held in an optical lattice, produced by the interference of several laser beams, can create the entanglement of many thousands of atoms, but only as a result of global operations performed jointly on all atoms<sup>9</sup>. Manipulation and detection of single atoms within the lattice are not yet possible, because the lattice spacings are too narrow for any optical beam to interact with individual atoms.

Technological improvements, stimulated by the rapidly developing field of quantum information processing<sup>10</sup>, will doubtless overcome these limitations in time. Brighter photon sources are being developed along with high-efficiency detectors with single-photon resolution. Continuous-variable approaches offer further possibilities beyond single-photon architectures. In addition, chip-based quantum circuits have been tested that contain silica-on-silicon waveguides<sup>11</sup>. Chip-based ion traps are also a promising route to engineering multipartite ion entanglement<sup>7</sup>, and advances in single-atom manipulation within optical lattices<sup>9</sup> could provide controllable multipartite

entanglement of thousands of particles.

The increased diversity of experimentally 'tamed' entangled particles brings theoretical challenges. The number of parameters needed to describe a state increases exponentially with the number of particles in the state, but this increased complexity also increases the potential utility of multipartite entangled states. For example, point-to-point quantum communication is more efficient when exploiting multipartite networks<sup>12</sup>. It is also possible that the computational power of multipartite states can be optimized or adapted to given architectures by making explicit use of specific entanglement properties such as intricate long-range correlations<sup>13</sup>.

We are only beginning to understand and exploit the power of multipartite entanglement. Current developments may seem to be tiny steps, but they could soon add up to a (quantum) leap not only in information science but also in our fundamental understanding of macroscopic quantum systems. ■ Markus Aspelmeyer is at the Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, 1090 Vienna, Austria. Jens Eisert is in the Department of Physics, University of Potsdam, 14469 Potsdam, Germany, and Imperial College London, UK. e-mails: markus.aspelmeyer@quantum.at; jense@qipc.org

1. Wieczorek, W. *et al.* *Phys. Rev. Lett.* **101**, 010503 (2008).
2. Greenberger, D. M., Horne, M. A., Shimony, A. & Zeilinger, A. *Am. J. Phys.* **58**, 1131–1143 (1990).
3. Raussendorf, R. & Briegel, H. J. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
4. Leibfried, D. *et al.* *Science* **304**, 1476–1478 (2004).
5. Pan, J.-W., Chen, Z.-B., Zukowski, M., Weinfurter, H. & Zeilinger, A. preprint at <http://arxiv.org/0805.2853> (2008).
6. Verstraete, F., Dehaene, J., De Moor, B. & Verschelde, H. *Phys. Rev. A* **65**, 052112 (2002).
7. Blatt, R. & Wineland, D. *Nature* **453**, 1008–1015 (2008).
8. Lu, C.-Y. *et al.* *Nature Phys.* **3**, 91–95 (2007).
9. Bloch, I. *Nature* **453**, 1016–1022 (2008).
10. Walmsley, I. A. *Science* **319**, 1211–1213 (2008).
11. Politi, A., Cryan, M. J., Rarity, J. G., Yu, S. & O'Brien, J. L. *Science* **320**, 646–649 (2008).
12. Acín, A., Cirac, J. I. & Lewenstein, M. *Nature Phys.* **3**, 256–259 (2007).
13. Gross, D. & Eisert, J. *Phys. Rev. Lett.* **98**, 220503 (2007).
14. Eisert, J. & Gross, D. preprint at <http://arxiv.org/abs/quant-ph/0505149> (2005).
15. Plenio, M. B. & Virmani, S. *Quant. Inf. Comp.* **7**, 001–051 (2007).

## Correction

The News & Views article "Materials science: A desirable wind up" (*Nature* **454**, 591–592; 2008), by Neil Mathur, considered work by D. Lebeugle *et al.* describing investigation of the multiferroic and magnetoelectric properties of single crystals of BiFeO<sub>3</sub> (*Phys. Rev. Lett.* **100**, 227602; 2008). Coverage of closely related work by V. Kiryukhin and colleagues (S. Lee *et al.* *Appl. Phys. Lett.* **92**, 192906; 2008), published just before that by Lebeugle *et al.*, was inadvertently omitted from the article.



producing new hindgut cells, albeit more slowly than during metamorphosis. The signal that triggers proliferation of the hindgut cells is not known. It is possible that dying larval cells send a signal to these cells instructing them to replenish hindgut cells during metamorphosis, in a process similar to the apoptotic mechanism that triggers homeostasis in the fly midgut<sup>4</sup> and the mammalian intestine<sup>8</sup>.

The concerted action of signalling molecules that regulate communication between the niche and the gut epithelium is crucial for maintaining the balance between cell renewal and differentiation. In the mouse gut, interplay between the signalling molecules Wnt, Hedgehog, BMP and Notch determines whether stem cells self-renew or differentiate. Wnt signalling is activated in the crypt and maintains cells in a proliferative state; increased activity of the Wnt pathway leads to enlarged crypts — often causing intestinal tumours<sup>2</sup> — whereas on its inhibition, crypts disappear. The signalling pathway mediated by Notch acts jointly with Wnt to sustain stem-cell proliferation, and is essential for the differentiation of specific cell types. Hedgehog signalling promotes differentiation and restricts crypt formation. This is accomplished, at least in part, through its effect on BMP signalling<sup>1</sup>.

It is fascinating to discover that the same pathways regulate cell proliferation in the fly hindgut. Takashima *et al.*<sup>6</sup> find that Wingless (Wg) — the fly protein related to Wnt — is expressed in a narrow stripe of cells at the anterior of the hindgut proliferation zone (Fig. 1). The outcome of perturbing Wg signalling is remarkably similar to that observed following perturbations of the Wnt signal in the mammalian intestine: whereas increased expression of the *wg* gene leads to expansion of the proliferation zone (reminiscent of crypt enlargement), blocking this gene's expression shrinks the pool of dividing cells (reminiscent of crypt disappearance). Furthermore, the Hedgehog pathway regulates the proliferative behaviour of these cells in *Drosophila* as it does in mammals; when this pathway is blocked, cells do not differentiate, remaining in a proliferative state.

Takashima and colleagues' work clearly demonstrates that the mechanism of cell proliferation in the adult fly hindgut follows on from that at work during development of this tissue. Whether the process in adults corresponds to a distinct homeostatic mechanism or is a continuation of development remains to be determined. Another question is why there is a need for fresh adult hindgut cells. In the case of the adult fly, midgut stem cells are required to replenish dying intestinal cells<sup>4</sup>, and the same is true for the mammalian intestine<sup>1,8</sup>. What is the mechanism of cell loss in the hindgut? Does shedding or apoptosis also occur there? In *Drosophila* we have an ideal model to address these questions, thus furthering our understanding of intestinal homeostasis. ■

Chrysoula Pitsouli and Norbert Perrimon are in the Department of Genetics, Harvard Medical

School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.  
e-mail: perrimon@receptor.med.harvard.edu

1. Crosnier, C., Stamatakis, D. & Lewis, J. *Nature Rev. Genet.* **7**, 349–359 (2006).
2. Reya, T. & Clevers, H. *Nature* **434**, 843–850 (2005).
3. Micchelli, C. A. & Perrimon, N. *Nature* **439**, 475–479 (2006).

4. Ohlstein, B. & Spradling, A. *Nature* **439**, 470–474 (2006).
5. Ohlstein, B. & Spradling, A. *Science* **315**, 988–992 (2007).
6. Takashima, S., Mkrtchyan, M., Younossi-Hartenstein, A., Merriam, J. R. & Hartenstein, V. *Nature* **454**, 651–655 (2008).
7. Skaer, H. in *The Development of Drosophila melanogaster* (ed. Bate, M.) 941–1012 (Cold Spring Harbor Laboratory Press, 1993).
8. Hall, P. *et al. J. Cell Sci.* **107**, 3569–3577 (1994).

## STRUCTURAL BIOLOGY

# It's not all in the family

Baruch I. Kanner

**There are no sequential snapshots of a transporter protein as it mediates the simultaneous passage of ions and solutes into a cell. Comparing different snapshots of structurally related transporters offers fascinating insights.**

Ion-coupled transporters are molecular machines that allow the passage of specific solutes, such as nutrients and neurotransmitters, across the cell membrane by transferring them together with one or more cations, often sodium ions. The energy released as these cations move into the cell down their concentration gradient is used to power the 'uphill' movement of the solute. For proper function, the transporter, which is embedded in the cell membrane, must alternately expose a binding site at either side of the membrane, capturing its cargo on one side and releasing it on the other — a principle called alternating access. In recent years, the structures of several ion-coupled transporters have been resolved, but in only a single conformation during transport. If we are to understand the molecular mechanism underlying transport, we need to know the structure of the transporter in different conformations. Faham *et al.*<sup>1</sup> take an initial, rather unexpected, step towards this goal by documenting in *Science* the first crystal structure of the sodium-coupled galactose transporter vSGLT.

A bacterial protein, vSGLT is a member of the family of solute sodium symporters (SSS). Members of this family have crucial roles in human health, and mutations in glucose and iodide symporters result in metabolic disorders. These symporters share no significant similarity in amino-acid sequence to members of another family — neurotransmitter sodium symporters (NSS). Yet Faham and colleagues find that there is an unexpected structural resemblance between vSGLT and LeuT (ref. 2), a bacterial member of the NSS family.

Although both vSGLT and LeuT crystallize as dimers<sup>1,2</sup>, in each case the monomers seem to be the functional units. Each monomer has a single binding pocket located at the interface of two inverted structural repeats (transmembrane  $\alpha$ -helices TM2–TM6 and TM7–TM11 in vSGLT, and TM1–TM5 and TM6–TM10 in LeuT). Another striking feature common to the two transporters is that, in each structure,

two transmembrane domains — TM2 and TM7 in vSGLT, and TM1 and TM6 in LeuT — have an interruption in the  $\alpha$ -helix. Such interruptions, which were first observed in the transmembrane domains of the calcium pump<sup>3</sup>, expose oxygen and nitrogen atoms in the peptide bonds of the amino-acid chain, for direct interaction with the solute and its co-transported cations.

Numerous functional studies on different types of transporter have found indirect evidence for alternating access, one of the most suggestive analyses being that of lactose permease of the bacterium *Escherichia coli*<sup>4</sup>. But what is particularly exciting about the structure of LeuT, and now Faham and colleagues' structure of vSGLT, is that whereas the former captures LeuT with its binding site facing the extracellular milieu, the opposite is true for vSGLT. So a comparison of these two structures allowed the authors<sup>1</sup> to provide the first direct clue for alternating access in membrane transporters. The two structures apparently result from structural rearrangements of several transmembrane  $\alpha$ -helices (Fig. 1, overleaf).

Unlike the vSGLT structure, that of LeuT is of high enough resolution to allow the identification of bound sodium ions ( $\text{Na}^+$ ). Nevertheless, Faham *et al.*<sup>1</sup> predict the existence of such a  $\text{Na}^+$ -binding site in vSGLT — and in other members of the SSS family — at a similar position to one of the  $\text{Na}^+$ -binding sites of LeuT (referred to as Na2). They obtain experimental support for this by studying the effects of mutations in an evolutionarily conserved serine amino-acid residue on TM9. Although the conservation of a  $\text{Na}^+$ -binding site gives extra confidence that the two structures can be compared for mechanistic purposes, it will be essential to obtain structures of the outward- and inward-facing conformations of a single transporter.

As for the position of galactose in the binding pocket of vSGLT, the authors could identify it unambiguously. This sugar seems to be sandwiched between groups of hydrophobic residues. The intracellular-exit pathway appears



### 50 YEARS AGO

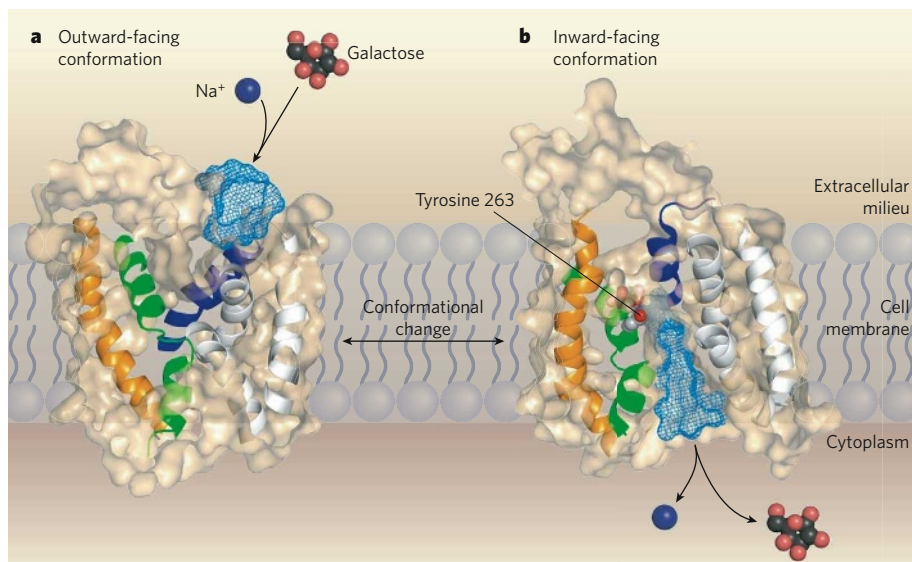
It is well known that various organisms can be cooled to very low temperatures in liquid gases without injury, provided that they have previously been desiccated to some extent. Extracellular freezing is considered to be a mode of dehydrating living cells; and some organisms, at least frost-resistant ones, may withstand freezing of this type even at very low temperatures ... Our experiments show that, after sufficient extracellular freezing, an intact insect can be kept alive at an extremely low temperature without any antifreeze agent. We used a 'slug caterpillar', *Cnidocampa flavescens* (Walk.) in the overwintering prepupal stage ... [T]he larvae in their cocoons were first frozen in a special refrigerator, in which the air temperature was lowered from  $-5^{\circ}$  to  $-90^{\circ}\text{C}$ . in  $1\frac{1}{2}$  hr. and maintained at that temperature for 45 min. The insects were then rewarmed in air at room temperature. After thawing, about one-third of the sixty larvae revived.

From *Nature* 2 August 1958.

### 100 YEARS AGO

Mr C. Kenrick Gibbons has presented to the Zoological Gardens a large number of the small fresh-water fish from Barbados known as "millions" (*Girardinus poecilloides*). These little fish, which have been placed in a tank in the tortoise house, are of special interest because of their supposed action in preventing malaria. Malaria is very much less common in Barbados than in other West Indian Islands, and it has been suggested that this freedom is due to the presence of enormous quantities of the "millions" in the fresh-water pools. The little fish are very voracious, and destroy large numbers of the larvae of mosquitoes that spread malaria ... It is understood that experiments are going to be made with the introduction of these fish into tropical countries where malaria is prevalent.

From *Nature* 30 July 1908.



**Figure 1 | The entrance and exit routes through the vSGLT transporter.** Faham *et al.*<sup>1</sup> find that the core (transmembrane  $\alpha$ -helices (TM) 2–11) of the vSGLT structure, which captures this transporter in a conformation facing the inside of the cell, is similar to the core (TM1–TM10) of the LeuT transporter, which faces the outside. A comparison of the two structures allowed the authors to identify the transmembrane helices (orange, green and blue) that undergo structural rearrangements when the protein switches from (a) facing outwards, taking up galactose and a sodium ion ( $\text{Na}^+$ ), to (b) facing inwards, allowing ion and solute release into the cell. Helices that undergo little movement are shown as white. Tyrosine amino acid 263 is indicated, and extracellular (a) and intracellular (b) vestibules are shown as a blue mesh. (Adapted from ref. 1.)

as a large hydrophilic vestibule that is separated from the binding pocket by the tyrosine amino-acid residue 263, and the displacement of this residue leads to the release of galactose into the cytoplasm (Fig. 1b). On the other side, the extracellular gate consists of a considerable protein mass that extends from the solute-binding site to the extracellular surface. The opposite situation is encountered in the outward-facing LeuT structure (Fig. 1a and ref. 2), where just a few residues separate the binding pocket and the extracellular vestibule, whereas the intracellular gate is very substantial<sup>2</sup>.

How does alternating access work? Clearly, it must involve substrate and cation binding as a trigger for the conformational change<sup>1</sup>. It has been proposed that, during transport, either the two interrupted transmembrane helices themselves<sup>2</sup>, or a bundle of these helices plus additional transmembrane domains<sup>5</sup>, move as a unit relative to the rest of the protein. The latter model<sup>5</sup> is based on the internal symmetry of the LeuT structure, which is also present in members of other transporter families. In this model, it is proposed that the conformational change leading to the opening of the intracellular-access pathway simultaneously closes its extracellular counterpart. Specifically, it is predicted that the intracellular-access pathway of LeuT is lined by the intracellular parts of TM1, TM5, TM6 and TM8. Studies on another member of the NSS family, SERT, which is the transporter for the neurotransmitter serotonin, reveal<sup>5</sup> that these very transmembrane domains have access to the aqueous intracellular medium, under conditions where the transporter is expected to face the inside of the cell. Such observations

not only support this elegant model<sup>5</sup> but are also compatible with the vSGLT structure<sup>1</sup>.

A different model<sup>6</sup> is based on the observation that LeuT has a second solute-binding site located at its extracellular vestibule, which was previously shown<sup>7,8</sup> to also accommodate antidepressants. The basic idea here is that the second substrate's binding to the external site results in the opening of the internal gate.

Clearly, additional functional and structural studies are required to shed light on the mechanism of alternating access. For example, after the cargo has been released into the internal compartment, the empty transporter reorients its binding pocket to start a new cycle. Moreover, the mere binding of sodium to the transporter seems to induce conformational changes that enable the transporter to bind the solute. To understand the mechanism of ion-coupled transport, it will be crucial to obtain structural information on the additional conformations of not only transporters of the NSS and SSS families, but also members of other transporter families. ■

Baruch I. Kanner is in the Department of Biochemistry, Hebrew University Hadassah Medical School, Jerusalem 91120, Israel.  
e-mail: kannerb@cc.huji.ac.il

1. Faham, S. *et al.* *Science* doi:10.1126/science.1160406 (2008).
2. Yamashita, A. *et al.* *Nature* **437**, 215–223 (2005).
3. Toyoshima, C. *et al.* *Nature* **405**, 647–650 (2000).
4. Kaback, H. R. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 491–494 (2007).
5. Forrest, L. R. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 10338–10343 (2008).
6. Shi, L. *et al.* *Mol. Cell* **30**, 667–677 (2008).
7. Singh, S. K., Yamashita, A. & Goux, E. *Nature* **448**, 952–956 (2007).
8. Zhou, Z. *et al.* *Science* **317**, 1390–1393 (2007).



# Does Rft1 flip an *N*-glycan lipid precursor?

Arising from: J. Helenius *et al.* *Nature* **415**, 447–450 (2002)

Protein *N*-glycosylation requires flipping of the glycolipid  $\text{Man}_5\text{GlcNAc}_2$ -diphosphate dolichol ( $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol) across the endoplasmic reticulum (ER)<sup>1–3</sup>. Helenius *et al.*<sup>4</sup> report genetic evidence suggesting that Rft1, an essential ER membrane protein in yeast, is required directly to translocate  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol. We now show that a specific ER protein(s), but not Rft1, is required to flip  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol in reconstituted vesicles. Rft1 may have a critical accessory role in translocating  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol *in vivo*, but the  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flippase itself remains to be identified.

We assayed flipping (Fig. 1a) by exploiting the organic-solvent-resistant interaction between  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol and the lectin concanavalin A (Con A)<sup>5</sup>. We reconstituted unilamellar, egg phosphatidylcholine vesicles containing tritium-labelled  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol. Organic solvents extracted ~50% of the tritium-labelled lipid from vesicles treated with Con A. This corresponds to the pool of tritiated  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol located in the inner leaflet; that in the outer leaflet was captured by Con A and not extracted. When we tested proteoliposomes containing Triton X-100-extracted ER membrane proteins (a source of  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flippase), capture of [<sup>3</sup>H] $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol approached ~100%, consistent with flipping of the glycolipid from the inner leaflet to the Con A-accessible outer leaflet. Thus, capture above 50% provides a measure of flipping.

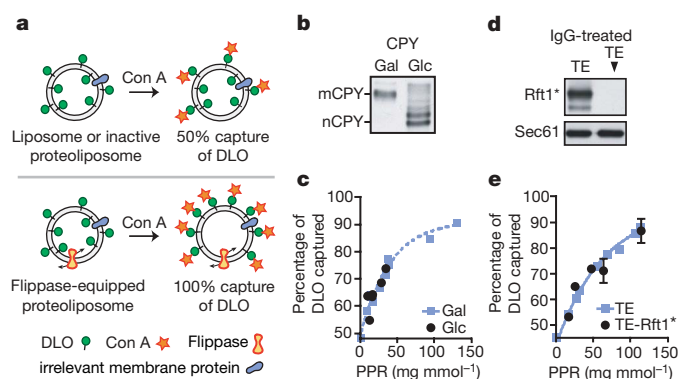
By reconstituting vesicles with different protein amounts, we generated dose-response plots (for example, Fig. 1c, e) with an initial slope that is proportional to the abundance of the flippase in the mixture of Triton X-100-extracted ER membrane proteins (~0.3% (w/w))<sup>6–8</sup>. As in other ER flipping events<sup>6–11</sup>,  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flipping was rapid, ATP-independent and protease-sensitive<sup>12</sup>. The reconstituted activity was substrate-specific:  $\text{Man}_7\text{GlcNAc}_2$ -PP-Dol was flipped over 10-fold more slowly than  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol, and glycerophospholipids were not flipped at all<sup>12</sup>.

To determine whether Rft1 flips  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol, we tested Triton X-100-extracted ER membrane proteins from Rft1-depleted yeast. We used strain YG1137 where Rft1 expression is controlled by

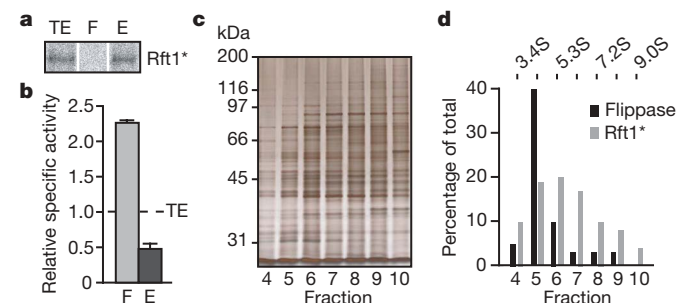
the glucose-repressible *GAL1-10* promoter<sup>4</sup>. As described<sup>4</sup>, *N*-glycoproteins are hypoglycosylated in glucose-grown YG1137 (Fig. 1b). Identical dose-response plots were obtained on reconstituting Triton X-100-extracted ER membrane proteins from galactose- and glucose-grown YG1137 (Fig. 1c), indicating that flippase abundance in the two extracts is independent of Rft1. Triton X-100-extracted ER membrane proteins were also prepared from YCF40 cells that require Protein A-tagged Rft1 (Rft1\*) for viability. These extracts were treated with IgG beads to deplete Rft1\* specifically and quantitatively (Fig. 1d). Reconstitution and assay of mock-treated and IgG-treated membrane proteins revealed identical dose-response plots (Fig. 1e), indicating that Rft1 does not contribute to the  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flippase activity of Triton X-100-extracted ER membrane proteins.

We chromatographed Triton X-100-extracted ER membrane proteins from YCF40 cells on CM-Sepharose and eluted the bound proteins with salt. Although Rft1\* was quantitatively recovered in the eluate (Fig. 2a), the specific activity of this fraction was low (Fig. 2b). However, the flow-through fraction lacked Rft1\* but had high specific activity. We also fractionated Triton X-100-extracted ER membrane proteins by velocity gradient sedimentation. Flippase activity peaked sharply at ~4.2S (3.8–4.5S in independent experiments) whereas rapidly sedimenting (>5.5S) proteins were inactive (Fig. 2c, d). Rft1\* sedimented broadly (2.5–9S) and its presence did not correlate with activity (Fig. 2d). We conclude that  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flippase activity is due to a specific ER membrane protein(s) in the ~4.2S fraction, but not to Rft1.

Although the phenotype of Rft1-depleted cells<sup>4</sup> indicates that Rft1 has a role in flipping  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol *in vivo*, our biochemical data indicate that it is not the flippase itself. Thus, Rft1 may act analogously to Lec35/MPDU1 (refs 13, 14), a protein whose proposed role in the utilization of dolichol-phosphomonosaccharides in the ER is evident *in vivo* but not *in vitro*<sup>13</sup>. One possibility is that in living cells Rft1 controls transfer of  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol from a biosynthetic compartment within the ER to the flippase. Compartmentalization would not exist in proteoliposomes or microsomes, obviating the requirement for Rft1 *in vitro*. Indeed, sealed microsomes from Rft1-depleted YG1137 cells convert *in situ* synthesized



**Figure 1 | Biochemical tests of the role of Rft1 in flipping  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol.** **a**, Flippase assay in reconstituted vesicles ([<sup>3</sup>H] $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol is indicated as dolichol-linked oligosaccharide (DLO)). **b**, **c**, Flippase activity in Triton X-100-extracted ER membrane proteins (TE) from Rft1-depleted cells. **b**, immunoblot of carboxypeptidase Y (CPY) from galactose (Gal)- and glucose (Glc)-grown YG1137 cells (mCPY, mature CPY; nCPY, non-glycosylated CPY). **c**, Flippase activity of reconstituted TEs from Glc- and Gal-grown YG1137 cells. **d**, **e**, Flippase activity of TE lacking Rft1. YCF40-TE was incubated with IgG-resin to immunodeplete Rft1\* (Protein A-tagged Rft1). **d**, Immunoblots of Rft1\* and Sec61 (control protein) in mock-treated and IgG-treated TE. **e**, Control and immunodepleted TEs were reconstituted and assayed for flippase activity.



**Figure 2 | Biochemical separation of Rft1 and  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flippase activity.** **a**, **b**, An Rft1-enriched fraction has low  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flippase activity. YCF40-TE was fractionated on CM-Sepharose. **a**, Immunoblot of Rft1\* in the TE, flow-through (F) and salt-eluted (E) fractions. TE, the F and E fractions, were reconstituted and assayed for flippase activity (**b**). Specific activity (relative to TE) was calculated as described<sup>6–9</sup> and reflects flippase abundance in the preparation. **c**, **d**, Fractionation of TE by velocity gradient sedimentation<sup>8</sup> enriches  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flippase activity ~4-fold and resolves it from Rft1\*. **c**, Silver-stained SDS-PAGE gel of fraction equivalents. **d**, Recovery of  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol flippase activity and Rft1\* in the fractions.

GlcNAc<sub>2</sub>-PP-Dol to Man<sub>5</sub>GlcNAc<sub>2</sub>-PP-Dol (J.S.R. and C.J.W., unpublished observations), a process that requires flipping of Man<sub>5</sub>GlcNAc<sub>2</sub>-PP-Dol<sup>2,3,5</sup>.

## METHODS

Yeast strain YCF40 (haploid; same strain background as YG1137<sup>4</sup>) lacks genomic *RFT1* but contains pJH19 (provided by J. Helenius and M. Aebi), a multicopy plasmid expressing Rft1\* from the *STT3* promoter. [<sup>3</sup>H]Man<sub>5</sub>GlcNAc<sub>2</sub>-PP-Dol was prepared from [2-<sup>3</sup>H]mannose-labelled *Δalg3Δalg5* cells. Membrane protein preparation and proteoliposome reconstitution were as described<sup>6–8</sup>. To assay flipping, vesicle aliquots were incubated at 4 °C for 30 min with buffer alone, with Con A, or with Con A and detergent, and were then treated with yeast mannan and extracted with solvent<sup>12</sup>. The distribution of [<sup>3</sup>H]Man<sub>5</sub>GlcNAc<sub>2</sub>-PP-Dol between the extract and Con A-containing pellet was determined for each sample. The 'buffer alone' and '+Con A+detergent' samples were used to calibrate the results.

**Christian G. Frank**<sup>†</sup>, **Sumana Sanyal**<sup>1</sup>, **Jeffrey S. Rush**<sup>2</sup>,  
**Charles J. Waechter**<sup>2</sup> & **Anant K. Menon**<sup>1</sup>

<sup>1</sup>Department of Biochemistry, Weill Cornell Medical College, New York, New York 10065, USA.

e-mail: akm2003@med.cornell.edu

<sup>2</sup>Department of Molecular and Cellular Biochemistry, University of Kentucky College of Medicine, Lexington, Kentucky 40536, USA.

<sup>†</sup>Present address: Fundación Caubet-CIMERA, 07110 Bunyola, Mallorca, Illes Balears, Spain.

Received 6 November 2007; accepted 3 June 2008.

1. Pomorski, T. & Menon, A. K. Lipid flippases and their biological functions. *Cell. Mol. Life Sci.* **63**, 2908–2921 (2006).

2. Schenk, B., Fernandez, F. & Waechter, C. J. The ins(ide) and out(side) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiol.* **11**, 61R–70R (2001).
3. Helenius, J. & Aebi, M. Transmembrane movement of dolichol linked carbohydrates during N-glycoprotein biosynthesis in the endoplasmic reticulum. *Semin. Cell Dev. Biol.* **13**, 171–178 (2002).
4. Helenius, J. *et al.* Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415**, 447–450 (2002).
5. Snider, M. D. & Rogers, O. C. Transmembrane movement of oligosaccharide-lipids during glycoprotein synthesis. *Cell* **36**, 753–761 (1984).
6. Menon, A. K., Watkins, W. E. & Hrafnisdóttir, S. Specific proteins are required to translocate phosphatidylcholine bidirectionally across the endoplasmic reticulum. *Curr. Biol.* **10**, 241–252 (2000).
7. Chang, Q. L., Gummadi, S. N. & Menon, A. K. Chemical modification identifies two populations of glycerophospholipid flippase in rat liver ER. *Biochemistry* **43**, 10710–10718 (2004).
8. Vehring, S. *et al.* Flip-flop of fluorescently labeled phospholipids in proteoliposomes reconstituted with *Saccharomyces cerevisiae* microsomal proteins. *Eukaryot. Cell* **6**, 1625–1634 (2007).
9. Vishwakarma, R. A. & Menon, A. K. Flip-flop of glycosylphosphatidylinositols (GPIs) across the ER. *Chem. Commun.* **2005**, 453–455 (2005).
10. Rush, J. S. & Waechter, C. J. Transmembrane movement of a water-soluble analogue of mannosylphosphoryldolichol is mediated by an endoplasmic reticulum protein. *J. Cell Biol.* **130**, 529–536 (1995).
11. Rush, J. S. & Waechter, C. J. Transbilayer movement of Glc-P-dolichol and its function as a glucosyl donor: protein-mediated transport of a water-soluble analog into sealed ER vesicles from pig brain. *Glycobiol.* **8**, 1195–1205 (1998).
12. Sanyal, S., Frank, C. G. & Menon, A. K. Distinct flippases translocate glycerophospholipids and oligosaccharide diphosphate dolichols across the endoplasmic reticulum. *Biochemistry* (in the press).
13. Anand, M. *et al.* Requirement of the *Lec35* gene for all known classes of monosaccharide-P-dolichol-dependent glycosyltransferase reactions in mammals. *Mol. Biol. Cell* **12**, 487–501 (2001).
14. Schenk, B. *et al.* MPDU1 mutations underlie a novel human congenital disorder of glycosylation, designated type If. *J. Clin. Invest.* **108**, 1687–1695 (2001).

doi:10.1038/nature07165

# Helenius *et al.* reply

Replying to: Frank, C. G., Sanyal, S., Rush, J. S., Waechter C. J. & Menon A. K. *Nature* **454**, doi:10.1038/nature07165 (2008)

The controlled translocation of polar phospholipid head groups between the two leaflets of biological membranes is thought to be mediated by proteins called flippases. Several candidate polyisoprene phosphate–oligosaccharide flippases have been described<sup>1–3</sup>, including Rft1 for Dol-PP-GlcNAc<sub>2</sub>Man<sub>5</sub> by us<sup>4</sup> and now questioned by Frank *et al.*<sup>5</sup>.

Flippase activities remain difficult to measure directly, and although biochemical approaches have been developed to monitor the flipping of different phospholipids<sup>6</sup>, these have failed to identify any flippases. There are difficulties in addressing the topological orientation of lipids in a membrane and the purification of flippases because these depend on the use of detergents that perturb native membrane structures. Although all flippase assays reveal a protein-dependent lipid flipping, only one was linked to a genetically identified flippase, Wzx<sup>7</sup>. However, that approach did not confirm the genetically defined flipping of the same glycolipid by two other Wzx flippases.

On the basis of a new biochemical assay, Frank *et al.*<sup>5</sup> fail to correlate the flipping of Dol-PP-GlcNAc<sub>2</sub>Man<sub>5</sub> with the presence of Rft1, our genetically identified flippase<sup>4</sup>. They conclude that Rft1 does not represent the flippase. However, they do not provide evidence that the measured flippase activity reflects the *in vivo* situation. Others have found<sup>8</sup> that using detergent to prepare membrane extracts generates denatured membrane proteins<sup>8</sup>, which can provide continuity between the leaflets of the membranes and thereby serve as artefactual conduits for polar moieties. The protein dependence of the reported activities and the frequently observed loss of substrate specificity of flippase activities *in vitro* are in accordance with this, which would explain the difficulties in correlating flippase activity with the presence of specific proteins.

More reliable assays are needed to characterize flippase activities. The purification and functional reconstitution of a homogenous flippase call for specificity controls and supporting genetic evidence. The caveats imposed by *in vitro* experimentation alone are just too daunting.

**Jonne Helenius**<sup>1†</sup>, **Davis T. W. Ng**<sup>2†</sup>, **Cristina L. Marolda**<sup>3</sup>, **Peter Walter**<sup>4</sup>,  
**Miguel A. Valvano**<sup>3</sup> & **Markus Aebi**<sup>1</sup>

<sup>1</sup>Institute of Microbiology, Swiss Federal Institute of Technology, Zürich, CH-8092 Zürich, Switzerland.

e-mail: aebi@micro.biol.ethz.ch

<sup>2</sup>Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

<sup>3</sup>Department of Microbiology and Immunology, University of Western Ontario, London, Ontario N6A 5C1, Canada.

<sup>4</sup>Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0448, USA.

<sup>†</sup>Present addresses: Biotechnology Centre, University of Technology Dresden, D-01307 Dresden, Germany (J.H.); Temasek Life Sciences Laboratory and Department of Biological Sciences, National University of Singapore, Singapore 117604 (D.T.W.N.).

1. Alaimo, C. *et al.* Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25**, 967–976 (2006).
2. Raetz, C. R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635–700 (2002).
3. Yan, A., Guan, Z. & Raetz, C. R. An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in *Escherichia coli*. *J. Biol. Chem.* **282**, 36077–36089 (2007).
4. Helenius, J. *et al.* Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415**, 447–450 (2002).



5. Frank, C. G. Sanyal, S. Rush, J. S., Waechter C. J. & Menon A. K. Does Rft1 flip an *N*-glycan lipid precursor? *Nature* **454**, doi:10.1038/nature07165 (2008).
6. Pomorski, T. & Menon, A. K. Lipid flippases and their biological functions. *Cell. Mol. Life Sci.* **63**, 2908–2921 (2006).
7. Rick, P. D. *et al.* Evidence that the *wzx* gene of *Escherichia coli* K-12 encodes a protein involved in the transbilayer movement of a trisaccharide-lipid intermediate in the assembly of enterobacterial common antigen. *J. Biol. Chem.* **278**, 16534–16542 (2003).
8. Kol, M. A. *et al.* Translocation of phospholipids is facilitated by a subset of membrane-spanning proteins of the bacterial cytoplasmic membrane. *J. Biol. Chem.* **278**, 24586–24593 (2003).

doi:10.1038/nature07164

# A new class of homoserine lactone quorum-sensing signals

Amy L. Schaefer<sup>1</sup>, E. P. Greenberg<sup>1</sup>, Colin M. Oliver<sup>2</sup>, Yasuhiro Oda<sup>1</sup>, Jean J. Huang<sup>1</sup>, Gili Bittan-Banin<sup>1</sup>, Caroline M. Peres<sup>3</sup>, Silke Schmidt<sup>4</sup>, Katarina Juhaszova<sup>1</sup>, Janice R. Sufrin<sup>2</sup> & Caroline S. Harwood<sup>1</sup>

**Quorum sensing is a term used to describe cell-to-cell communication that allows cell-density-dependent gene expression. Many bacteria use acyl-homoserine lactone (acyl-HSL) synthases to generate fatty acyl-HSL quorum-sensing signals, which function with signal receptors to control expression of specific genes. The fatty acyl group is derived from fatty acid biosynthesis and provides signal specificity, but the variety of signals is limited. Here we show that the photosynthetic bacterium *Rhodopseudomonas palustris* uses an acyl-HSL synthase to produce *p*-coumaroyl-HSL by using environmental *p*-coumaric acid rather than fatty acids from cellular pools. The bacterium has a signal receptor with homology to fatty acyl-HSL receptors that responds to *p*-coumaroyl-HSL to regulate global gene expression. We also found that *p*-coumaroyl-HSL is made by other bacteria including *Bradyrhizobium* sp. and *Silicibacter pomeroyi*. This discovery extends the range of possibilities for acyl-HSL quorum sensing and raises fundamental questions about quorum sensing within the context of environmental signalling.**

Dozens of species of Proteobacteria use members of the LuxI family of signal generators to produce fatty acyl-HSL signals to which co-evolved members of the LuxR family of transcription factors respond. The lengths of the fatty acyl groups have been reported to vary from 4 carbons to 18 carbons depending on the specific system. LuxIR-type systems influence a variety of processes in bacteria including virulence in some animal and plant pathogens<sup>1,2</sup>. The anoxygenic phototrophic soil bacterium *Rhodopseudomonas palustris* CGA009 has a chromosomal luxIR-type pair, *rpaI* (*rpa0320*) and *rpaR* (*rpa0321*)<sup>3</sup> (Fig. 1a and Supplementary Fig. 1). Recently, we found that *rpaI* expression is activated specifically by growth of *R. palustris* on *p*-coumarate<sup>4</sup>, a major aromatic monomer of lignin polymers, which comprise over 30% of all plant dry material. *R. palustris* degrades a wide variety of aromatic compounds<sup>3</sup>. To investigate the relationship between quorum sensing and *p*-coumarate in *R. palustris*, we sought to identify the product of RpaI activity. We first used a set of bioassays<sup>5</sup>, which will detect all known acyl-HSLs. There was no detectable response to extracts of culture fluid from *R. palustris* grown on *p*-coumarate in any of the assays. Thus, we hypothesized that the *R. palustris* quorum-sensing signal has a unique structure, and only the cognate receptor RpaR will respond to the RpaI-generated signal.

## A bioassay for the *R. palustris* signal

To test our hypothesis that *R. palustris* produces a unique acyl-HSL quorum-sensing signal, we required an RpaR-dependent promoter. Because luxI-type genes are often positively autoregulated by their cognate R protein<sup>1,2</sup> and the R proteins often bind to identifiable inverted repeat sequences called lux box-like sequences<sup>6</sup>, we examined the DNA sequence upstream of the *rpaI* open reading frame and found two lux box-like elements centred at -76 and -35 bp upstream of the *rpaI* ATG start codon (Fig. 1a, b). We generated *R. palustris* CGA814, which has a chromosomal *rpaI::lacZ* mutation to test activation of the *rpaI* promoter by solvent extracts of *R. palustris*

wild-type cultures grown on succinate in the presence and absence of *p*-coumarate. Extracts from cultures containing *p*-coumarate, but not succinate only, activated *rpaI::lacZ* expression (Fig. 1c). To assess whether *p*-coumarate was the most potent inducer of this system, we surveyed extracts of *R. palustris* grown on succinate plus each of 14 additional aromatic compounds (Supplementary Fig. 2) for *rpaI* promoter induction. Extracts from *p*-coumarate-grown cultures elicited the strongest response, although plant-derived aromatic compounds with modest hydroxylation differences on the phenyl ring exhibited small (<10%) amounts of induction (Fig. 1c). We also tested the ability of a number of synthetic fatty acyl-HSLs to activate the *rpaI* promoter and all results were negative (see Methods Summary).

## Signal purification and identification

Culture extracts were separated by C<sub>18</sub>-reverse-phase high-performance liquid chromatography (HPLC). By using the *rpaI::lacZ* strain as a reporter we detected a single peak of activity in *p*-coumarate-grown cultures (Fig. 2a). By using high-resolution mass spectrometry we calculated a mass of 247.0844 for the compound in the active fraction, which is in agreement with a molecular formula of C<sub>13</sub>H<sub>13</sub>NO<sub>4</sub>. This formula is inconsistent with any fatty acyl-HSL, but it corresponds to *p*-coumaroyl-HSL (hereafter called *pC*-HSL). Thus, we synthesized *pC*-HSL and monitored the ability of the synthetic compound to activate the *rpaI::lacZ* reporter (Fig. 2b). The expression of *rpaI::lacZ* was activated by synthetic *pC*-HSL and the half-maximal response was 5–10 nM, a sensitivity similar to those reported for fatty acyl-HSL signal receptors<sup>7</sup>. The response curve for the natural compound (Fig. 2c) was similar to that of synthetic *pC*-HSL. Stationary phase cultures of *R. palustris* grown on *p*-coumarate under either anaerobic photoheterotrophic or aerobic chemoheterotrophic conditions typically had activity equivalent to about 1 μM *pC*-HSL. These extracellular levels are comparable to signal concentrations produced by many bacteria with fatty acyl-HSL signalling systems<sup>7</sup>.

<sup>1</sup>Department of Microbiology, University of Washington, Washington 98195, USA. <sup>2</sup>Molecular Pharmacology and Cancer Therapeutics Program, Roswell Park Cancer Institute, State University of New York at Buffalo, Buffalo, New York 14263, USA. <sup>3</sup>Danisco Genencor, Palo Alto, California 94304, USA. <sup>4</sup>Institute of Molecular Biosciences, University of Frankfurt, Frankfurt 60438, Germany.



Tandem mass spectrometry of the natural product (Fig. 2d) showed a pattern indistinguishable from that of synthetic *p*C-HSL (Fig. 2e). Synthetic *p*C-HSL exhibited an HPLC elution profile similar to the natural product in both a methanol gradient (Fig. 2a) and in a 20:80 methanol:water isocratic separation, as well as a similar absorption spectrum (broad peak of maximum absorbance at 306 nm and a minor peak at 224 nm, data not shown). An NMR analysis of the natural and synthetic compounds (Supplementary Fig. 3) supports our conclusion that they are both *p*-coumaroyl-HSL (Fig. 2f). Furthermore, the NMR analysis indicates that the double bond in the aryl side chain is in the *trans* configuration for both the natural and synthetic compounds. We have not characterized the signal produced by *R. palustris* when we add *m*- or *o*-coumarate to cultures. As a rule, acyl-HSL synthases show some lack of specificity with respect to the acyl substrate and have limited, but measurable, abilities to use acyl substrates with similarity to the natural substrate<sup>8–10</sup>. Thus, it is

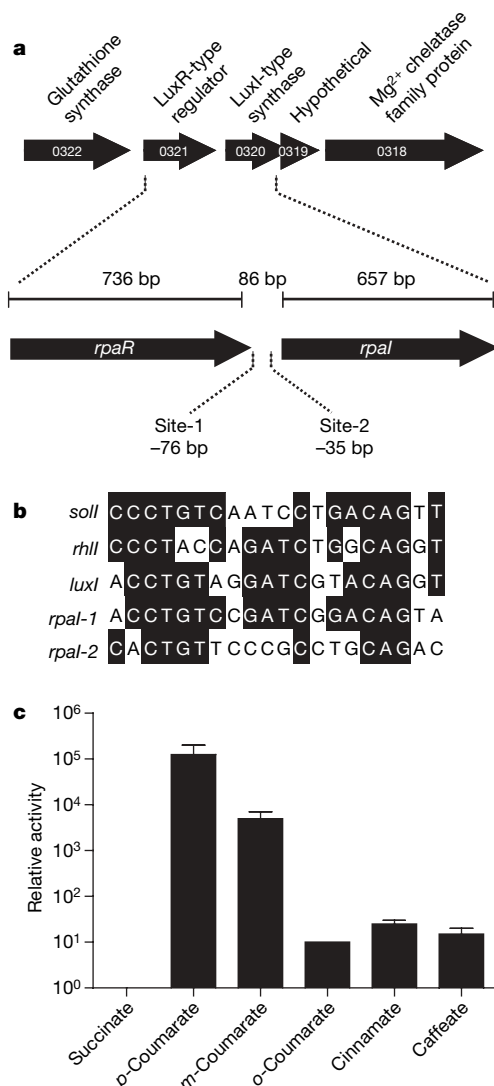
reasonable to assume that when fed *m*- or *o*-coumarate, *R. palustris* synthesizes *m*- or *o*-coumaroyl-HSL, but this requires further investigation.

### The *p*C-HSL-dependent regulon

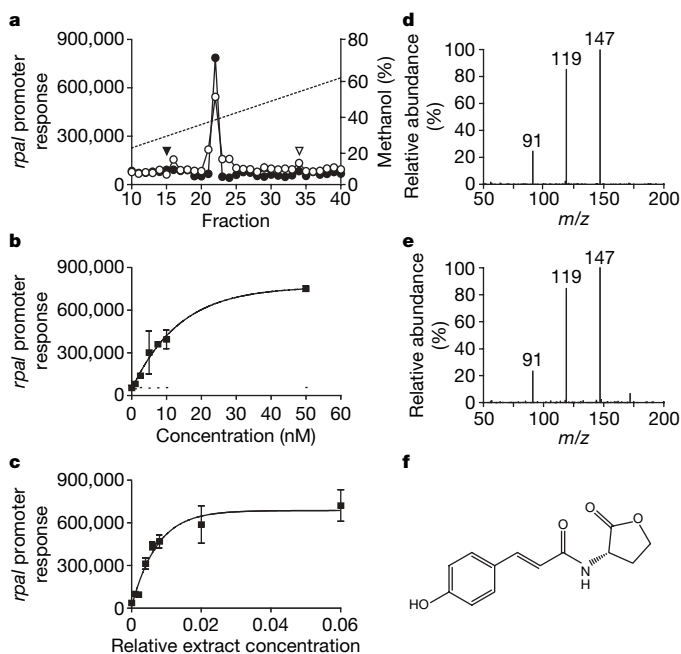
A defining characteristic of fatty acyl-HSL signals is their ability to control expression of specific sets of genes<sup>1,2</sup>. We wanted to test whether the novel aryl-HSL produced by *R. palustris* served to regulate genes in addition to the gene for its own production. Thus, we performed transcriptome analyses of bacteria grown on succinate, with or without added *p*C-HSL. Expression of 17 genes was altered more than 2.5-fold ( $P < 0.001$ ) by *p*C-HSL addition (Supplementary Table 1). As expected, expression of *rpaI* and the adjacent *rpa0319* (predicted to comprise an operon)<sup>11</sup> was induced by *p*C-HSL. Over one-half of the regulated genes have no known function, but several genes are annotated as encoding chemotaxis functions (*rpa1096*, *rpa1674*, *rpa3185* and *rpa4684*). None of the genes thought to be involved in *p*-coumarate degradation<sup>4</sup> is controlled by *p*C-HSL, consistent with our observation that the *R. palustris* *rpaI* mutant grows normally on *p*-coumarate (data not shown). As predicted, all of the *p*C-HSL-controlled genes were also regulated in *R. palustris* grown on *p*-coumarate (where *p*C-HSL is self generated) but not on succinate (a condition where *p*C-HSL is not made)<sup>4</sup>. These data suggest that *p*C-HSL is indeed a signal that regulates gene transcription in *R. palustris*.

### RpaR-RpaI form a quorum-sensing circuit

To assess whether the *p*C-HSL signal functions in a quorum-sensing-dependent manner, we examined its accumulation during *R. palustris* growth in the presence of *p*-coumarate (Fig. 3a). Freshly inoculated



**Figure 1 | The *rpaI* gene has potential RpaR binding sites in its promoter and plant-derived aromatic acids activate its expression.** **a**, The *rpaR-rpaI* gene region with the locations of two potential *lux* box-like elements shown below. Annotations of genes are shown above. **b**, Comparison of the elements with known *lux* box-like elements from *Ralstonia solanacearum* *soll*, *Pseudomonas aeruginosa* *rhII* and the *Vibrio fischeri* *luxI*. **c**, Relative signal activity extracted from cultures grown with 10 mM succinate or succinate plus 0.5 mM *p*-coumarate, *m*-coumarate, *o*-coumarate, cinnamate, or caffeate. Signal activity was measured with the *R. palustris* *rpaI::lacZ* reporter as described in the text. Relative  $\beta$ -galactosidase activity was calculated from duplicate cultures. Error bars indicate the range.



**Figure 2 | Evidence indicating that the *R. palustris* quorum-sensing signal is *p*C-HSL.** **a**, HPLC analysis of culture fluid extracts from *p*-coumarate-grown *R. palustris* (filled circles) and synthetic *p*C-HSL (open circles). 3OC6-HSL (filled triangle) and 3OC8-HSL (open triangle) were eluted in the indicated fractions. The dashed line shows the methanol gradient. **b**, **c**, The *rpaI* promoter response to synthetic *p*C-HSL (**b**) and *p*-coumarate-grown wild-type culture fluid extract (**c**). **d**, **e**, MS/MS daughter scans of the *m/z* 248 (*M*+*H*) ion from the natural product (**d**) and synthetic *p*C-HSL (**e**). **f**, The proposed *R. palustris* quorum-sensing signal, *p*C-HSL. Duplicate samples (**b**, **c**) were analysed. Error bars indicate the range. The *rpaI* promoter response was measured using the *R. palustris* CGA814 *rpaI::lacZ* chromosomal fusion.  $\beta$ -Galactosidase activity was measured with a luminescence reagent and is expressed as relative luminescence units.

cultures grew without lag, whereas *pC*-HSL production did not increase until later in growth, at which point aryl-HSL levels increased rapidly. This phenomenon is a characteristic of most quorum-sensing systems<sup>12,13</sup>, which are positively autoregulated. By using real-time polymerase chain reaction (PCR), we examined *rpaI* expression during growth. Expression increased in concordance with *pC*-HSL concentration (89-fold increase over the initial 7 h of growth). These results are consistent with the conclusion that *pC*-HSL functions as a quorum-sensing signal.

With fatty acyl-HSL quorum-sensing systems, the *luxR* linked to the *luxI* homologue codes for the cognate signal-receptor transcription factor. Thus, one would assume that *rpaR* codes for a *pC*-HSL receptor. To test the assumption that RpaR is required for *pC*-HSL-dependent gene expression, we created reporter plasmids containing an *rpaI::lacZ* fusion with or without *rpaR* for use in the heterologous host, *Pseudomonas aeruginosa* MW1. When *rpaR* was present, *rpaI::lacZ* expression increased as a function of *pC*-HSL concentration (black bars, Fig. 3b). Without *rpaR*, *rpaI::lacZ* expression was high regardless of aryl-HSL levels (white bars, Fig. 3b). This is consistent with the hypothesis that RpaR functions as a *pC*-HSL-responsive repressor of the *rpaI* promoter. Thus, RpaR function is analogous to that of the LuxR homologue, EsaR, from *Pantoea stewartii*<sup>14</sup>, which is a fatty acyl-HSL-responsive repressor protein.

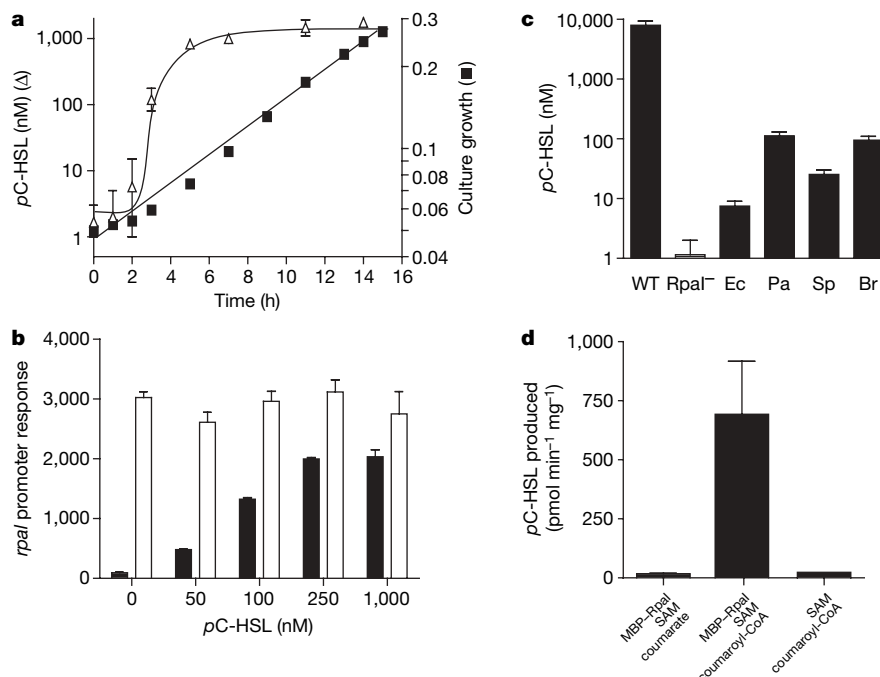
### Purified RpaI is an aryl-HSL synthase

We next investigated whether *pC*-HSL is generated by the *rpaI* gene product. We first showed that extracts from a *p*-coumarate-grown *rpaI* mutant contained no detectable *pC*-HSL (Fig. 3c). This is consistent with our hypothesis that *pC*-HSL is generated by RpaI. To obtain further evidence we expressed *rpaI* in both *Escherichia coli*, which cannot use *p*-coumarate as a carbon source, and *P. aeruginosa*, which can use *p*-coumarate as a carbon source. In either species *rpaI* directed the synthesis of *pC*-HSL, but only when exogenous *p*-coumarate was provided (Fig. 3c). Production of *pC*-HSL in recombinant *E. coli* and *P. aeruginosa* was low compared to production of

*pC*-HSL by *R. palustris*. There are several possible explanations for this finding. For example, levels of active RpaI might be low in the recombinant bacteria or the recombinant bacteria might have a limited ability to activate *p*-coumarate.

To demonstrate definitively its role as a *pC*-HSL synthase we purified RpaI as a maltose binding protein (MBP) fusion for use in *in vitro* aryl-HSL synthesis experiments. Our results show that MBP-RpaI can catalyse the synthesis of *pC*-HSL from *S*-adenosylmethionine (SAM) and *p*-coumaroyl-coenzyme A (CoA) (Fig. 3d). This is similar to other LuxI-type enzymes that use SAM as the homoserine lactone ring donor and a thioesterified acyl group for the side-chain substrate<sup>8–10</sup>. Fatty acyl-HSL synthases prefer acyl-acyl carrier protein (ACP) intermediates from fatty acid biosynthesis, but can use acyl-CoA compounds to a lesser degree<sup>8</sup>. *p*-Coumaroyl-CoA formation has been detected in *R. palustris* crude cell extracts<sup>15</sup>, consistent with reports from other bacteria suggesting that the initial step in *p*-coumarate degradation is aryl-CoA formation<sup>16</sup>. We are unaware of any report describing a *p*-coumaroyl-aryl carrier protein (ArCP) metabolite, although there is precedence for ArCPs in bacteria. For example, the *E. coli* EntB carboxy-terminal domain functions as an ArCP for 2,3-dihydroxybenzoic acid during enterobactin siderophore synthesis<sup>17</sup>. Interestingly, the *R. palustris* gene *rpa0319*, which is immediately downstream of *rpaI* and the expression of which is activated by *pC*-HSL addition (Supplementary Table 1), codes for a small ACP-sized polypeptide with a serine that could serve as a site for pantetheinylation. This polypeptide might function as an aryl donor for RpaI. We also note there are three genes in the *R. palustris* genome that are annotated as ACPs. Any one of these could function as an aryl carrier. Thus, it seems possible that a *p*-coumaroyl-ArCP might serve as the natural substrate for RpaI. However, regardless of which *p*-coumaroyl-thioester (-CoA or -ArCP) is the preferred substrate, our experiments demonstrate that RpaI is a *pC*-HSL synthase.

The *p*-coumarate requirement for signal production, even in heterologous hosts, suggests that exogenous *p*-coumarate, most likely derived from plants, serves as the source of the coumaroyl moiety in



**Figure 3** | *pC*-HSL is made during growth, requires RpaR for quorum-sensing-dependent activity, is synthesized by RpaI, and is made by other bacteria. **a**, *pC*-HSL production (open triangles) during growth (filled squares) of *R. palustris* with *p*-coumarate. Culture growth was measured as optical density at 660 nm. **b**, *pC*-HSL-dependent *rpaI* promoter activity requires RpaR. Expression of *rpaI::lacZ* from a plasmid with (black bars) and

without (white bars) *rpaR* in *P. aeruginosa* MW1. **c**, *pC*-HSL production by *R. palustris* wild type (WT), an *rpaI* mutant (RpaI<sup>-</sup>), *E. coli* DH5α (pRpaI) (Ec), *P. aeruginosa* MW1 (pRpaI) (Pa), *S. pomeroi* DSS-3 (Sp) and *Bradyrhizobium* BTAi1 (Br) when grown with *p*-coumarate. **d**, *pC*-HSL synthesis by purified MBP-RpaI. Values are averages of triplicates (a) or duplicates (b–d). Error bars indicate the range.



*pC*-HSL. We found that *p*-coumarate at concentrations as low as 1  $\mu$ M supported *pC*-HSL production in *R. palustris* cultures grown with succinate as the main carbon source. Production of the *Agrobacterium tumefaciens* fatty acyl-HSL also requires chemicals provided by another organism. However, for *A. tumefaciens* the chemical is a plant-produced opine required to activate expression of the LuxR-type regulator<sup>18</sup>, not to serve as part of the actual quorum-sensing signal.

### Other bacteria produce *pC*-HSL

Is *pC*-HSL synthesis unique to *R. palustris*? To address this question we screened for *pC*-HSL activity in extracts of bacteria that can use *p*-coumarate as a carbon source and that are known from genomic sequence analyses to have *luxI* homologues. When grown in the presence of *p*-coumarate, two bacteria, *Bradyrhizobium* sp. BTAi1 and *Silicibacter pomeroyi* DSS-3, produced relatively small amounts of material that behaved biologically and chemically like *pC*-HSL (Fig. 3c). *Bradyrhizobium* BTAi1 is a photosynthetic, nitrogen-fixing symbiont that forms root and stem nodules on *Aeschynomene* plant species<sup>19</sup>. *S. pomeroyi* is a coastal bacterioplankton member of the *Roseobacter* clade, an abundant marine group of importance in global sulphur and carbon cycles<sup>20</sup>. Thus, *pC*-HSL production does not appear to be unique to *R. palustris*. That *Bradyrhizobium* and *S. pomeroyi* produce low levels of *R. palustris* signal activity suggests that the growth conditions we have used are not optimal for signal production, or the more interesting possibility that *p*-coumarate is similar to but not the natural substrate for their aryl-HSL synthases.

### Discussion

Our description of a non-fatty acyl-HSL (aryl-HSL) quorum-sensing signal that is derived from an exogenously supplied substrate expands the range of potential acyl-HSLs enormously. The concept that only straight-chain fatty acids in metabolic pools can serve as substrates for quorum-sensing signal production no longer holds and one can imagine any number of organic acids from the environment serving as the acyl group on a quorum-sensing signal. In the case we describe here the acyl group is derived from a plant metabolite, *p*-coumarate. The use of *p*-coumarate for quorum-sensing signal production results in a single signal that integrates two distinct cues: sufficiently high bacterial population densities and the availability of a particular exogenous substrate. This could be beneficial to an organism that controls functions via quorum sensing that are useful only under a particular set of conditions. *p*-Coumarate is produced by plants as a constituent of the major plant polymer lignin and *p*-coumarate synthesis is also stimulated by tissue damage and other stresses<sup>21,22</sup>. Because *pC*-HSL production requires a source of *p*-coumarate we speculate that there is an intimate relationship between *pC*-HSL-producing bacteria and specific plants and that this relationship involves *pC*-HSL signalling. It is distinctly possible that *pC*-HSL serves not only as an intraspecies bacterial signal but also as an inter-kingdom signal to a host plant. The discovery of *pC*-HSL opens up this question.

### METHODS SUMMARY

*Rhodopseudomonas palustris* strains CGA009<sup>3</sup> and CGA814 (*rpaI::lacZ*) were grown photoheterotrophically in photosynthetic medium (PM) with succinate (10 mM) plus *p*-coumarate (0.5–3 mM, where indicated) as described elsewhere<sup>23</sup>. When indicated, plant-derived aromatic acids (0.5 mM, Supplementary Fig. 2) were added to PM succinate cultures. *P. aeruginosa* MW-1 (*lasI*<sup>−</sup>, *rhlI*<sup>−</sup>)<sup>24</sup>, *E. coli* DH5 $\alpha$  and XL1-blue were grown in Luria broth (LB) plus 1 mM *p*-coumarate where indicated. *Bradyrhizobium* BTAi1 and *Silicibacter pomeroyi* DSS-3 were grown aerobically in PM medium and basal medium<sup>20</sup>, respectively, with 10 mM succinate and 1 mM *p*-coumarate. Bacteria were grown at 30 °C. The following synthetic fatty acyl-HSLs (1  $\mu$ M) were tested for their ability to activate the *rpaI* promoter as measured by  $\beta$ -galactosidase activity expressed from the *R. palustris* CGA814 *rpaI::lacZ* reporter strain: C4-HSL, 3OHC4-HSL, 3OHC6-HSL, 3OC6-HSL, C7-HSL, C8-HSL, 3OHC8-HSL, C12-HSL, 3OC12-HSL and C16-HSL. Plasmid and strain constructions are

described in the Methods. To monitor *pC*-HSL we mixed test samples (culture extracts, HPLC fractions, or synthetic compounds) with strain CGA814, incubated for 16 h in light, and then measured  $\beta$ -galactosidase activity<sup>24</sup>. Microarray analyses were a comparison of wild-type *R. palustris* CGA009 grown photoheterotrophically with succinate as the carbon source with and without 10  $\mu$ M synthetic *pC*-HSL. For purification of *p*-coumaroyl-HSL we grew cultures of *R. palustris* CGA009 with *p*-coumarate to early stationary phase, removed cells by centrifugation and then extracted signal from the culture fluid with ethyl acetate. The active material was purified by gradient and isocratic HPLC<sup>25</sup>. For synthesis of *pC*-HSL we developed a novel method involving microwave irradiation as described in detail in the Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 18 April; accepted 14 May 2008.**

**Published online 18 June 2008.**

- Fuqua, C. & Greenberg, E. P. Self perception in bacteria: quorum sensing with acylated homoserine lactones. *Curr. Opin. Microbiol.* **1**, 183–189 (1998).
- Waters, C. M. & Bassler, B. L. Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Dev. Biol.* **21**, 319–346 (2005).
- Larimer, F. W. et al. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nature Biotechnol.* **22**, 55–61 (2004).
- Pan, C. et al. Characterization of anaerobic catabolism of *p*-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics. *Mol. Cell. Proteomics* **7**, 938–948 (2008).
- Schaefer, A. L., Hanzelka, B. L., Parsek, M. R. & Greenberg, E. P. Detection, purification, and structural elucidation of the acylhomoserine lactone inducer of *Vibrio fischeri* luminescence and other related molecules. *Methods Enzymol.* **305**, 288–301 (2000).
- Devine, J. H., Shadel, G. S. & Baldwin, T. O. Identification of the operator of the *lux* regulon from the *Vibrio fischeri* strain ATCC7744. *Proc. Natl. Acad. Sci. USA* **86**, 5688–5692 (1989).
- Pearson, J. P. et al. Structure of the autoinducer required for expression of *Pseudomonas aeruginosa* virulence genes. *Proc. Natl. Acad. Sci. USA* **91**, 197–201 (1994).
- Parsek, M. R., Val, D. L., Hanzelka, B. L., Cronan, J. E. Jr & Greenberg, E. P. Acyl homoserine-lactone quorum-sensing signal generation. *Proc. Natl. Acad. Sci. USA* **96**, 4360–4365 (1999).
- Schaefer, A. L., Val, D. L., Hanzelka, B. L., Cronan, J. E. Jr & Greenberg, E. P. Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified *Vibrio fischeri* LuxI protein. *Proc. Natl. Acad. Sci. USA* **93**, 9505–9509 (1996).
- More, M. I. et al. Enzymatic synthesis of a quorum-sensing autoinducer through use of defined substrates. *Science* **272**, 1655–1658 (1996).
- Moreno-Hagelsieb, G. & Collado-Vides, J. A powerful non-homology method for the prediction of operons in prokaryotes. *Bioinformatics* **1** (suppl.) S329–S336 (2002).
- Hastings, J. W. & Greenberg, E. P. Quorum sensing: the explanation of a curious phenomenon reveals a common characteristic of bacteria. *J. Bacteriol.* **181**, 2667–2668 (1999).
- Nealson, K. H., Platt, T. & Hastings, J. W. Cellular control of the synthesis and activity of the bacterial luminescence system. *J. Bacteriol.* **104**, 313–322 (1970).
- Minogue, T. D., Wehland-von Trebra, M., Bernhard, F. & von Bodman, S. B. The autoregulatory role of EsaR, a quorum sensing regulator in *Pantoea stewartii* subsp. *stewartii*: evidence for a repressor function. *Mol. Microbiol.* **44**, 1625–1635 (2002).
- Harrison, F. H. *Peripheral Pathways of Anaerobic Benzoate Degradation in Rhodopseudomonas palustris*. PhD thesis, Univ. Iowa (2005).
- Peng, X., Misawa, N. & Harayama, S. Isolation and characterization of thermophilic Bacilli degrading cinnamic, 4-coumaric, and ferulic acids. *Appl. Environ. Microbiol.* **69**, 1417–1427 (2003).
- Gehring, A. M., Bradley, K. A. & Walsh, C. T. Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* **36**, 8495–8503 (1997).
- Fuqua, W. C. & Winans, S. C. A. LuxR-LuxI type regulatory system activates *Agrobacterium* Ti plasmid conjugal transfer in the presence of a plant tumor metabolite. *J. Bacteriol.* **176**, 2796–2806 (1994).
- Molouba, F. et al. Photosynthetic bradyrhizobia from *Aeschynomene* spp. are specific to stem-nodulated species and form a separate 16S ribosomal DNA restriction fragment length polymorphism group. *Appl. Environ. Microbiol.* **65**, 3084–3094 (1999).
- Buchan, A., Neidle, E. L. & Moran, M. A. Diverse organization of genes of the beta-ketoadipate pathway in members of the marine *Roseobacter* lineage. *Appl. Environ. Microbiol.* **70**, 1658–1668 (2004).
- Dixon, R. A. & Palva, N. L. Stress-induced phenylpropanoid metabolism. *Plant Cell* **7**, 1085–1097 (1995).
- Whetten, R. & Sederoff, R. Lignin biosynthesis. *Plant Cell* **7**, 1001–1013 (1995).
- Kim, M.-K. & Harwood, C. S. Regulation of benzoate-CoA ligase in *Rhodopseudomonas palustris*. *FEMS Microbiol. Lett.* **83**, 199–204 (1991).

24. Whiteley, M., Lee, K. M. & Greenberg, E. P. Identification of genes controlled by quorum sensing in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **96**, 13904–13909 (1999).
25. Singh, P. K. *et al.* Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **407**, 762–764 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank B. Howald and R. Lawrence for their expertise in MS analyses, and W. Tabaczynski for his expertise in NMR analysis. This work was supported by the US Army Research Office, US Department of Energy (C.S.H.), and the National Institute of General Medical Sciences (E.P.G.). C.M.O. was supported by a National Cancer Institute Training Grant.

**Author Contributions** A.L.S. performed all experiments except those described below. C.M.O. and J.R.S. performed and interpreted NMR analyses and synthesized pC-HSL. Y.O. assisted in substrate survey and growth curve experiments and performed RT-PCR analyses. J.J.H. constructed strains and performed experiments analysing RpaR-dependent gene activation. G.B.-B. and C.M.P. constructed strains. S.S. purified MBP-RpaI protein. K.J. synthesized p-coumaroyl-CoA. A.L.S., E.P.G. and C.S.H. designed experiments, analysed data and wrote the paper.

**Author Information** The primary microarray data have been deposited in the NCBI's Gene Expression Omnibus (GEO, [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)) under the accession number GSE10642. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to C.S.H. ([csh5@u.washington.edu](mailto:csh5@u.washington.edu)).



## METHODS

**Plasmid and strain constructions.** The *rpaI::lacZ* reporter strain, *R. palustris* CGA814, was constructed by replacing a 357-bp fragment of *rpaI* with a *lacZ::Km<sup>r</sup>* cassette<sup>26</sup> in pJQ200KS<sup>27</sup> and mating this construct into *R. palustris* CGA009. To investigate the role of *rpaR* in quorum-sensing regulation of *rpaI* expression, *lacZ* (with ribosomal binding site) from pHRP309<sup>26</sup> was cloned into pBBR1MCS-5<sup>28</sup> to create pBBR1MCS-5lacZ. The *rpaR*-containing *rpaI::lacZ* reporter plasmid, pRpaR-PrpaI, was constructed by amplifying a 1,040-bp DNA fragment containing 222-bp upstream of *rpaR*, *rpaR*, and the 86-bp intergenic region between *rpaR* and *rpaI* (Fig. 1a) from *R. palustris* and cloning it into pBBR1MCS-5lacZ. An *rpaI::lacZ* plasmid lacking *rpaR*, pPrpaI, was constructed by amplifying the 86-bp DNA region upstream of *rpaI* and cloning it into pBBR1MCS-5lacZ. To express RpaI in the heterologous hosts *E. coli* DH5 $\alpha$  and *P. aeruginosa* MW1, pPrpaI was constructed by amplifying the *rpaI* gene and cloning it into pBBR1MCS-5<sup>28</sup>.

**Detection of *p*-coumaroyl-HSL.** Cultures to be tested for *p*C-HSL were extracted twice with equal volumes of acidified ethyl acetate (EtAc, 0.1 ml glacial acetic acid per litre of solvent). Levels of *p*C-HSL were followed by using either the *R. palustris* strain CGA814 *rpaI::lacZ* reporter or *P. aeruginosa* MW1 (pRpaR-PrpaI). Test samples (culture extracts, HPLC fractions, or synthetic compounds) were added to 1.5 ml Eppendorf tubes (*R. palustris* CGA814 assay) or 13-mm glass vials (*P. aeruginosa* pRpaR-PrpaI assay) and the solvent was removed by evaporation under N<sub>2</sub>. *R. palustris* CGA814 in PM succinate (0.5 ml culture) was added to each test sample and incubated in light for 16 h. Cells were lysed with chloroform (10% vol/vol) and  $\beta$ -galactosidase activity was measured with a Tropix Galacto-Light Plus Kit (Applied Biosystems) as described previously<sup>24</sup>. Logarithmic-phase *P. aeruginosa* MW1 (pRpaR-PrpaI) cultures (0.5 ml) were added to each test sample. After 4 h shaking at 30 °C cells were lysed and  $\beta$ -galactosidase activity measured. For *p*C-HSL measurement we used a standard curve generated with synthetic *p*C-HSL. Results were similar regardless of whether the reporter strains were grown in the presence or absence of *p*-coumarate.

**Purification and identification of *p*-coumaroyl-HSL.** *R. palustris* CGA009 (4-l culture) was grown with 10 mM succinate and 3 mM *p*-coumarate to early stationary phase. Cell-free supernatant fluid was extracted twice with equal volumes of EtAc. The quorum-sensing signal was purified for MS from extracts as described previously<sup>25</sup> except that the final isocratic elution was in 20% methanol in water. In the 20% methanol gradient the natural product was eluted with a characteristic skewed absorbance peak (305 nm) and the synthetic *p*C-HSL showed a peak with the same shape and elution profile. MS/MS was performed with a Waters Micromass Quattro II (collision energy 30 eV, collision gas argon, cone 35 V). To obtain sufficient purified signal for NMR analysis, we extracted 13 l of cell culture. Before HPLC separation, the extracts were dried, dissolved in 50% methanol in water, and passed through a C<sub>18</sub> Sep-Pak cartridge (Waters Corp.) from which activity was eluted in 20–50% methanol.

**Chemical synthesis of *p*-coumaroyl-HSL.** PS-carbodiimide resin (0.5 g), *p*-coumarate (0.0382 g, predominantly trans, Sigma C9008), (S)-(-)- $\alpha$ -amino- $\gamma$ -butyrolactone HBr (0.0275 g) and 5 ml of dimethylformamide were added to a 10-ml reactival (Pierce). After microwave irradiation (1,000 W, 30 s, Milestone

START labstation), PS-carbodiimide resin was removed by vacuum filtration and solvents were removed *in vacuo*. The resulting residue was dissolved in acetonitrile and purified by preparative TLC (EtAc:Hex, 80:20 v/v). Product band ( $R_F = 0.14$ ) was eluted (ACN:MeOH, 80:20 v/v, 60 min) to yield a white solid (0.0095 g, 0.038 mmol, 24.8%). The final structure was confirmed by TLC (EtOAc:Hex, 80:20 v/v)  $R_F = 0.14$ , MS/MS (Fig. 2e) and <sup>1</sup>H-NMR (Supplementary Fig. 3).

**Gene expression analyses.** For transcriptome studies, wild-type *R. palustris* CGA009 was grown with succinate (conditions under which *p*C-HSL is not produced) to an absorbance of 0.1 at 660 nm. We then split the culture in half and added 10  $\mu$ M synthetic *p*C-HSL to one half. The amended and unamended cultures were further incubated until the density reached 0.5. Cellular RNA was isolated<sup>29</sup>, cDNA was prepared and transcriptome analysis was carried out with an Affymetrix *R. palustris* Custom GeneChip as previously described<sup>30</sup>. Real-time PCR reactions included 1 ng cDNA and 200 nM primers in 25  $\mu$ l of SYBR green PCR amplification master mix (Applied Biosystems). PCR conditions were 2 min at 50 °C, 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C, and 1 min at 60 °C. Genomic DNA was used as a standard and *fixJ* (*rpa4248*) constitutive expression was used as an internal control.

**Purification of MBP-RpaI fusion protein.** To obtain purified RpaI fusion protein, the *rpaI* gene was cloned into pMalc2 (New England Biolabs) creating pMBPrpaI. *E. coli* XL1-blue (pMBPrpaI) was grown at 30 °C in 1 l LB plus 2% glucose and ampicillin to an absorbance of 0.4 at 600 nm. MBP-RpaI expression was then induced (0.3 mM IPTG) and incubation continued at 16 °C overnight. Cells were pelleted, resuspended in a buffer<sup>9</sup> and broken by sonication. The MBP-RpaI fusion protein was purified from clarified cell extracts by amylose column chromatography (New England Biolabs protocol). We were unable to cleave MBP from RpaI using the factor X<sub>A</sub> cleavage site.

**In vitro *p*-coumaroyl-HSL synthesis assays.** Aryl-HSL synthesis assays were performed as reported previously<sup>8</sup> except that reaction mixtures (0.1 ml) contained 20 mM Tris pH 7.5, 10 mM MgCl<sub>2</sub>, 1 mM dithiothreitol, 4  $\mu$ g MBP-RpaI, 36 nCi carboxy-<sup>14</sup>C-labelled SAM (60 mCi mmol<sup>-1</sup>, final SAM concentration 60  $\mu$ M, American Radiochemical Company) and 40  $\mu$ M *p*-coumaroyl-CoA or *p*-coumarate and were incubated for 75 min at 30 °C. We synthesized *p*-coumaroyl-CoA as described previously<sup>15</sup>.

26. Parales, R. E. & Harwood, C. S. Construction and use of a new broad-host-range *lacZ* transcriptional fusion vector, pHRP309, for gram-bacteria. *Gene* **133**, 23–30 (1993).
27. Quandt, J. & Hynes, M. F. Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. *Gene* **127**, 15–21 (1993).
28. Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M. II & Peterson, K. M. pBBR1MCS: a broad-host-range cloning vector. *Biotechniques* **16**, 800–802 (1994).
29. Oda, Y. et al. Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodospseudomonas palustris*. *J. Bacteriol.* **187**, 7784–7794 (2005).
30. Rey, F. E., Heiniger, E. K. & Harwood, C. S. Redirection of metabolism for biological hydrogen production. *Appl. Environ. Microbiol.* **73**, 1665–1671 (2007).

## ARTICLES

# Switching on and off fear by distinct neuronal circuits

Cyril Herry<sup>1\*</sup>, Stephane Cioocchi<sup>1\*</sup>, Verena Senn<sup>1</sup>, Lynda Demmou<sup>1</sup>, Christian Müller<sup>1</sup> & Andreas Lüthi<sup>1</sup>

**Switching between exploratory and defensive behaviour is fundamental to survival of many animals, but how this transition is achieved by specific neuronal circuits is not known. Here, using the converse behavioural states of fear extinction and its context-dependent renewal as a model in mice, we show that bi-directional transitions between states of high and low fear are triggered by a rapid switch in the balance of activity between two distinct populations of basal amygdala neurons. These two populations are integrated into discrete neuronal circuits differentially connected with the hippocampus and the medial prefrontal cortex. Targeted and reversible neuronal inactivation of the basal amygdala prevents behavioural changes without affecting memory or expression of behaviour. Our findings indicate that switching between distinct behavioural states can be triggered by selective activation of specific neuronal circuits integrating sensory and contextual information. These observations provide a new framework for understanding context-dependent changes of fear behaviour.**

The amygdala is a key brain structure mediating defensive behaviour in states of fear and anxiety. Such states can be induced by classical auditory fear conditioning, in which an initially neutral auditory stimulus (the conditioned stimulus, CS) comes to elicit a fear response after pairing with an aversive foot shock (the unconditioned stimulus, US). Subsequent repetitive presentations of the CS alone induce a progressive decrease in the fear response, a phenomenon called extinction. Whereas firing of amygdala neurons is critical for the retrieval of conditioned fear memories<sup>1–6</sup>, their firing after the extinction of conditioned fear is thought to be constrained by local inhibitory circuits activated by the medial prefrontal cortex (mPFC)<sup>6–11</sup>. Converging evidence from animal studies indicates, however, that the basolateral complex of the amygdala (BLA), comprising the lateral (LA) and the basal (BA) nuclei, actively participates in fear extinction<sup>12–17</sup>. Although fear extinction is an active learning process eventually leading to the formation of a consolidated extinction memory<sup>16,17</sup>, it is a fragile behavioural state that is readily influenced by context<sup>18,19</sup>. Changing context results in the immediate recovery of the previously conditioned fear response, a process known as fear renewal<sup>18,19</sup>. *In vivo* pharmacological studies indicate that the hippocampus, which is reciprocally connected to the BLA<sup>20</sup>, processes contextual information during fear conditioning, extinction and renewal<sup>19,21–23</sup>. Thus, bi-directional changes in fear behaviour during extinction and context-dependent renewal are likely to be encoded within a distributed network containing the BLA, the mPFC and the hippocampus; however, the neuronal circuits mediating such behavioural transitions are not known. In particular, this raises the question of whether there are specialized circuits driving behavioural transitions in opposite directions.

To address this question, we used a combination of *in vivo* single-unit recordings and targeted pharmacological inactivation in behaving mice. Because the BA is strongly connected to the hippocampus<sup>20</sup> and to the mPFC<sup>24,25</sup>, and because extinction has previously been shown to induce the expression of the activity-dependent immediate early gene product Fos in BA neurons<sup>26</sup>, we focused our study on this sub-nucleus. Here we identify two distinct neuronal circuits differentially connected with the mPFC and the hippocampus, and show

that a rapid switch in the balance of activity between those circuits specifically drives behavioural transitions without being necessary for memory storage or behavioural expression.

## Distinct BA neurons encode fear and extinction

To examine plasticity of spike firing of individual BA neurons, C57Bl/6 mice were implanted with chronic recording electrodes and trained in a discriminative fear-conditioning paradigm (Fig. 1a). During training, mice learned to discriminate two auditory CS of different frequencies. One CS (the CS<sup>+</sup>) was paired with an aversive foot shock (US), whereas the second CS (CS<sup>−</sup>) was not paired. Twenty-four hours after fear conditioning, mice ( $n = 30$ ) exhibited a selective increase in fear behaviour (as measured by freezing) when exposed to the CS<sup>+</sup> in a different context (Fig. 1c). Extinction of conditioned fear behaviour was induced by exposing mice to 24 CS<sup>+</sup> presentations in the absence of any aversive stimuli. After extinction training, CS<sup>+</sup>-induced freezing behaviour was reduced back to pre-conditioning levels, and did not differ from CS<sup>−</sup>-induced freezing (Fig. 1c).

Analysis of changes in CS<sup>+</sup>- and CS<sup>−</sup>-evoked spike firing during extinction training revealed that BA neurons (259 recorded units; Fig. 1b) could be divided into distinct functional classes. Consistent with previous reports<sup>27,28</sup>, we found a class of neurons ('fear neurons';  $n = 43$  neurons, 22 mice; 17% of recorded units) that exhibited a selective increase in CS<sup>+</sup>-evoked spike firing during and after fear conditioning (Fig. 1d, Supplementary Fig. 1 and Supplementary Table 1). Subsequent extinction completely abolished this increase and converted it into a CS<sup>+</sup>-evoked inhibition (Fig. 1d). On average, spontaneous activity of fear neurons was not affected by fear conditioning or extinction (Supplementary Table 1). Thus, fear-conditioning-induced behavioural discrimination between the CS<sup>+</sup> and the CS<sup>−</sup>, and its reversal by extinction, was accurately reflected at the neuronal level by the discriminative and reversible activity of fear neurons.

During extinction training, another class of neurons emerged. In contrast to fear neurons, 'extinction neurons' ( $n = 35$  neurons, 20 mice; 14% of recorded units) did not show any increase in CS-evoked

<sup>1</sup>Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland.

\*These authors contributed equally to this work.



responses during or after fear conditioning, but instead showed a slight reduction (Fig. 1e). However, subsequent extinction training induced a marked and selective increase in CS<sup>+</sup>-evoked activity in these neurons (Fig. 1e), without any changes in spontaneous activity. Plotting extinction-induced changes in z-score for individual fear and extinction neurons revealed that the two populations were separated in a bi-modal distribution (Supplementary Fig. 2). The remaining neurons did not exhibit any changes in activity during extinction (Supplementary Table 1). Thus, changes in CS<sup>+</sup>-evoked firing of fear and extinction neurons were oppositely correlated with behavioural extinction.

Although these results demonstrate a specific activation of fear and extinction neurons by a given CS, they do not address the question of whether individual extinction neurons can function as fear neurons for another CS, or vice versa. We therefore trained mice in a discriminative

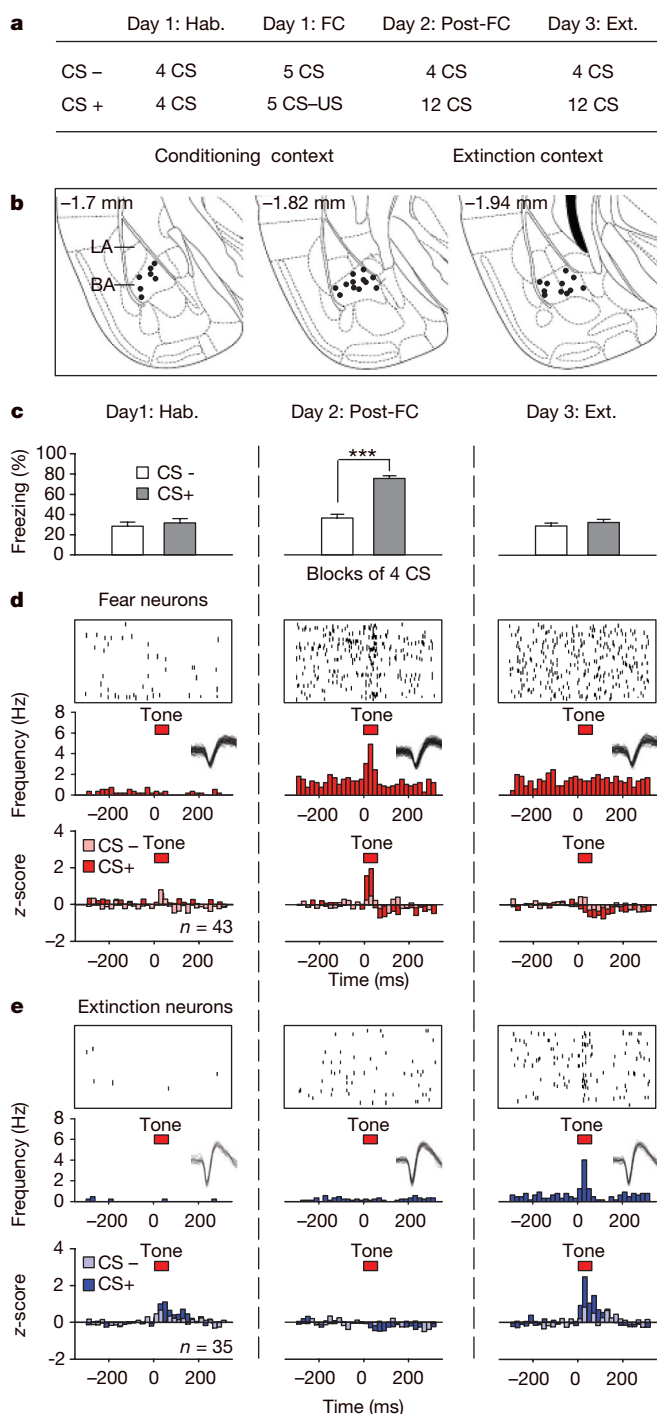
extinction paradigm (Fig. 2a). In this paradigm, two different CS (CS1 and CS2) were first fear-conditioned, followed by extinction of only one of them (CS1). At the end of extinction, mice exhibited selective freezing behaviour when exposed to the non-extinguished CS2 (Fig. 2b). Fear neurons and extinction neurons were identified during fear conditioning and extinction of CS1 according to the same criteria as described above, and CS1- and CS2-evoked spike firing was compared at the end of extinction. Whereas individual extinction neurons ( $n = 9$  neurons, 3 mice) responded to the extinguished CS (CS1), but not to the non-extinguished CS (CS2), fear neurons ( $n = 8$ , 3 mice) only fired during CS2 exposure, but remained unresponsive to the CS1 (Fig. 2c, d). These observations confirm that individual fear neurons and extinction neurons represent functionally distinct classes of neurons that can discriminate between extinguished and non-extinguished stimuli.

In addition to the BA, we also recorded from 38 neurons in the LA, which represents the main target of sensory afferents from the thalamus and cortex<sup>5</sup>. In keeping with previous studies<sup>1,29</sup>, we did not observe any LA neuron in which CS<sup>+</sup>-evoked firing increased during extinction. Although we cannot exclude the existence of such neurons in the LA, this may suggest that extinction neurons are specific for the BA, where they represent 14% of all recorded neurons.

### Activity balance predicts behaviour

Comparing the averaged time courses of CS-evoked activity of fear and extinction neurons during the acquisition of behavioural extinction indicated that significant behavioural changes occurred after the activity scores of the two populations of neurons crossed over (Fig. 3a). The largest changes in CS-evoked activity for both fear and extinction neurons occurred between the third and the fourth blocks of extinction training, which are separated by 24 h, suggesting that an overnight consolidation process may be required. To investigate further the exact time point during extinction learning at which fear and extinction neurons displayed a significant change in activity, we applied a change-point analysis algorithm<sup>30</sup>. Change-point analysis identifies the trial(s) exhibiting a significant change in neuronal activity or freezing behaviour relative to the preceding trials. This analysis confirmed that changes in neuronal activity precede behavioural changes, and revealed that the activity of extinction neurons started to increase one trial before the activity of fear neurons began to decline (Fig. 3b, c). Plotting activity changes of single fear and extinction neurons recorded in the same animal showed that the sequence of events is the same in an individual animal, and that such changes occur abruptly in an all-or-none manner (Fig. 3c). This is consistent with the idea that behavioural changes are driven by sequential switches in the activity of two distinct neuronal circuits.

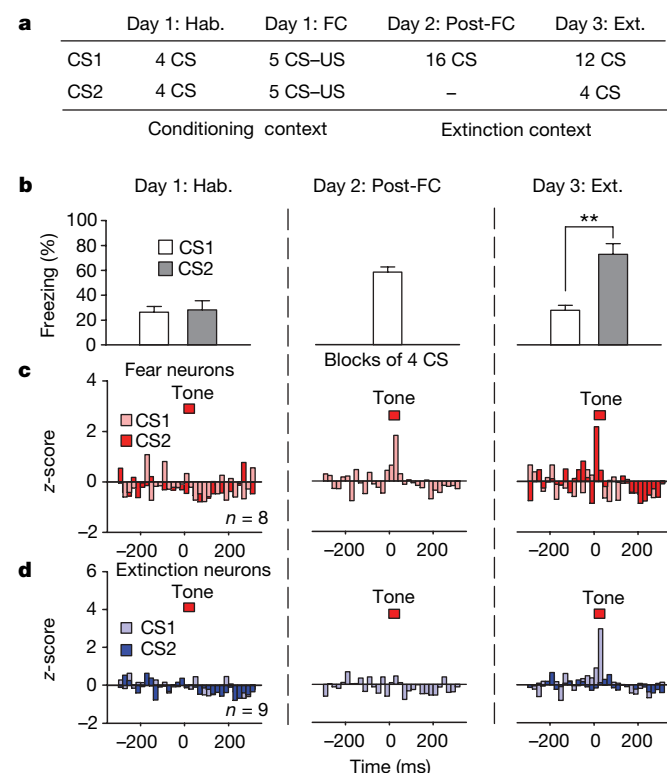
**Figure 1 | Distinct populations of BA neurons encode fear conditioning and extinction.** **a**, Experimental protocol. Ext., extinction; FC, fear conditioning; Hab., habituation. **b**, Coronal sections through the rostro-caudal extent of the amygdala, showing the location of the recording sites in the BA. **c**, Summary graphs illustrating behavioural data. During habituation, mice ( $n = 30$ ) exhibited equally low freezing levels in response to CS<sup>+</sup> and CS<sup>-</sup> exposure. Twenty-four hours after fear conditioning, presentation of the CS<sup>+</sup> (CS 1–4 on day 2), but not the CS<sup>-</sup>, evoked significantly higher freezing levels. After extinction, both CS<sup>+</sup> (CS 9–12 on day 3) and CS<sup>-</sup> elicited low freezing levels. Error bars indicate mean  $\pm$  s.e.m. **d**, **e**, Raster plots (top) and peristimulus time histograms (middle) illustrating selective changes in CS<sup>+</sup>-evoked firing of a representative fear (**d**) and an extinction (**e**) neuron. The duration of the auditory stimulus is indicated (red bar; tone). Insets show superimposed spike waveforms recorded during habituation, after fear conditioning and after extinction, respectively. Bottom: fear conditioning and extinction-induced changes in CS<sup>+</sup>-evoked firing of fear and extinction neurons. Fear neurons ( $n = 43$  neurons from 22 mice) exhibited a selective increase in CS<sup>+</sup>-evoked firing after fear conditioning ( $P < 0.001$  versus habituation or versus CS<sup>-</sup>), which was fully reversed on extinction. In contrast, CS<sup>+</sup>-evoked firing of extinction neurons ( $n = 35$  neurons from 20 mice) was selectively increased after extinction ( $P < 0.001$  versus post-FC or versus CS<sup>-</sup>). \*\*\* $P < 0.001$ .



### Rapid reversal of activity during fear renewal

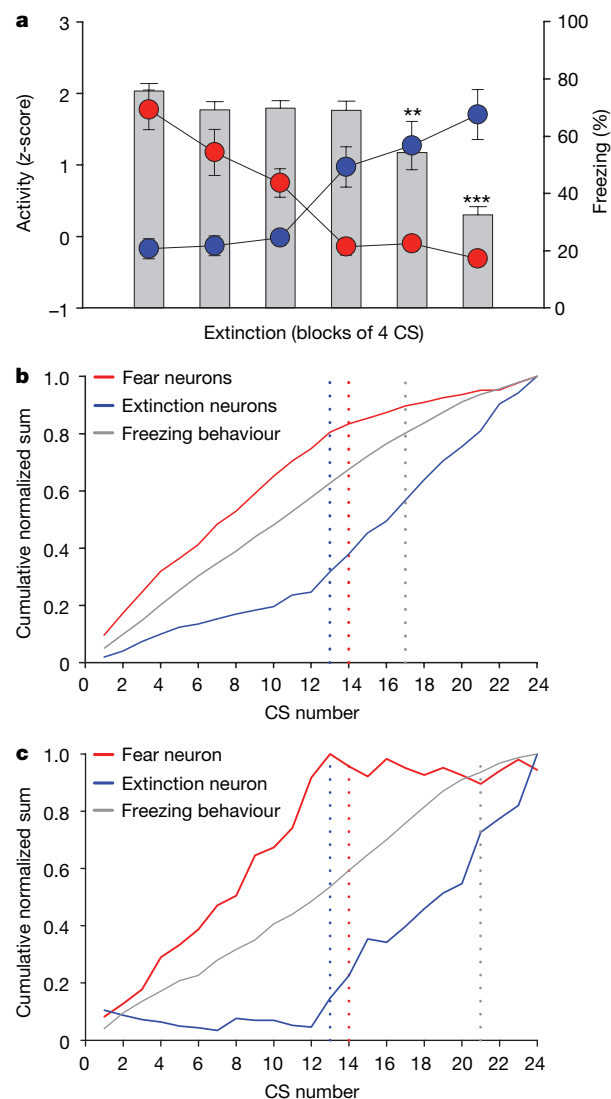
To test whether the activity of fear and extinction neurons represents the same behavioural values in a different paradigm, we analysed renewal of extinguished fear behaviour and associated changes in CS-evoked spike firing. To make sure that extinction memory was stably consolidated, mice ( $n = 15$ ) were tested for extinction memory 7 days after extinction training in the same context in which extinction training occurred (Fig. 4a). After successful recall of extinction memory (Fig. 4b), mice were transferred to the context in which they had been initially fear conditioned. Changing context resulted in a modest, but significant, increase in baseline freezing levels owing to contextual fear conditioning (Supplementary Fig. 3), and in a full renewal of the original cued fear memory (Fig. 4b).

During recall of extinction memory in the extinction context, presentation of the CS<sup>+</sup> induced a selective activation of extinction neurons ( $n = 14$ , 8 mice) with no effect on fear neurons ( $n = 19$ , 9 mice; Fig. 4c, d). Thus, activation of extinction neurons by an extinguished CS is not a transient phenomenon, but remains stable for at least one week. After placing the animals in the fear-conditioning



**Figure 2 | Fear and extinction neurons discriminate stimuli with different emotional significance.** **a**, Experimental design for discriminative extinction training. Initially, animals were fear conditioned to two distinct CS (CS1 and CS2). Both CS were paired with a US (CS-US). Subsequently, only one CS (CS1) was extinguished. **b**, Summary of behavioural data. During habituation, mice ( $n = 6$ ) exhibited equally low freezing levels in response to CS1 and CS2 exposure. After fear conditioning, presentation of the CS1 (CS 1–4) evoked significantly increased freezing levels. After extinction to CS1, CS1 exposure (CS 9–12) elicited low freezing levels, whereas CS2-evoked freezing behaviour remains high. Error bars indicate mean  $\pm$  s.e.m. **c**, Fear conditioning- and extinction-induced changes in CS1- and CS2-evoked firing of fear neurons ( $n = 8$  neurons from 3 mice). Twenty-four hours after fear conditioning (day 2), fear neurons exhibited increased firing in response to CS1 stimulation. After extinction of CS1, only CS2 stimulation elicited significant firing (day 3;  $P < 0.05$  versus CS1). **d**, Fear conditioning- and extinction-induced changes in CS1- and CS2-evoked firing of extinction neurons ( $n = 9$  neurons, 3 mice). After fear conditioning (day 2), extinction neurons did not respond to CS1 stimulation. After extinction of CS1, only CS1 stimulation elicited significant firing (day 3;  $P < 0.05$  versus CS2). \*\* $P < 0.01$ .

context, increased CS<sup>+</sup>-evoked freezing behaviour was associated with a complete reversal of spiking activity at the cellular level. Whereas extinction neurons stopped responding to CS<sup>+</sup> stimulation, fear neurons exhibited a significant and selective increase in CS<sup>+</sup>-evoked spike firing (Fig. 4d). Extinction-resistant neurons were not significantly activated during renewal (not shown). Thus, a switch in the balance of activity between fear and extinction neurons not only



**Figure 3 | Sequential switches in neuronal activity precede behavioural changes.** **a**, Averaged time courses of freezing behaviour (grey bars;  $n = 30$  mice) and neuronal activity (z-scores) of BA fear neurons (red circles;  $n = 43$ ) and extinction neurons (blue circles;  $n = 35$ ) during extinction training. Significant behavioural changes (that is, decreased freezing levels) occurred after activity scores of fear and extinction neurons have crossed over. Error bars indicate mean  $\pm$  s.e.m. **b**, Change-point analysis confirms that changes in neuronal activity preceded behavioural changes, and demonstrates that the activity of extinction neurons started to increase one trial before the activity of fear neurons changed. The plot represents the cumulative sums of the averaged and normalized z-scores of fear and extinction neurons, and freezing behaviour during extinction training. Change points are indicated by dotted lines. **c**, Normalized cumulative sums of the z-scores of a single fear neuron and a single extinction neuron recorded in the same animal together with the corresponding freezing behaviour during extinction training. Change-point analysis reveals that the extinction neuron abruptly switched on one trial before the fear neuron switched off. Changes in neuronal activity preceded behavioural changes. Change points are indicated by dotted lines. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .



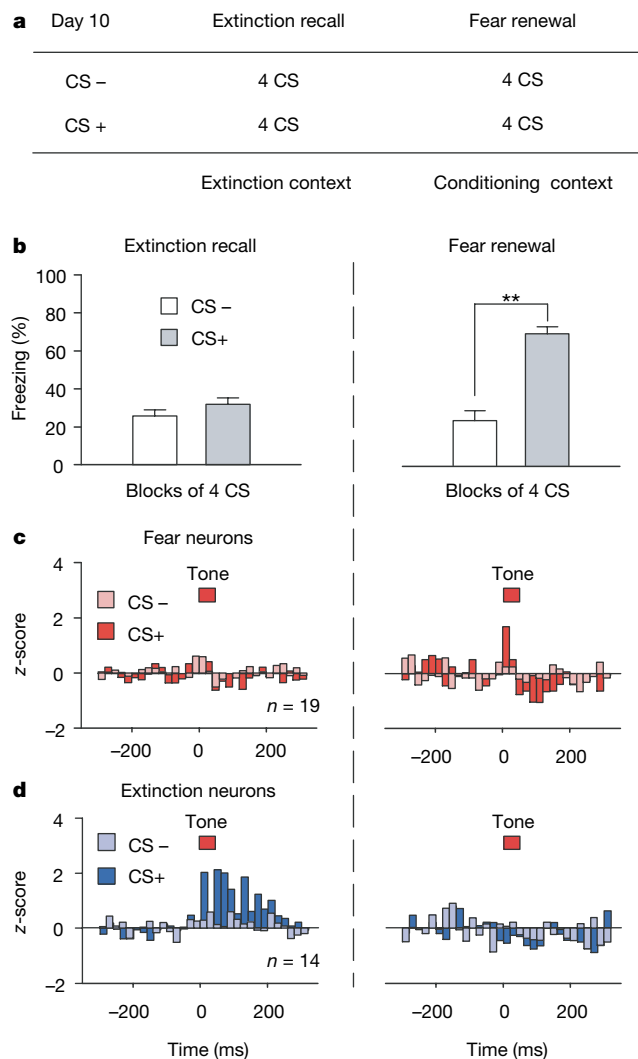
reflects extinction but also parallels rapid context-dependent renewal of conditioned fear responses.

### Differential long-range connectivity

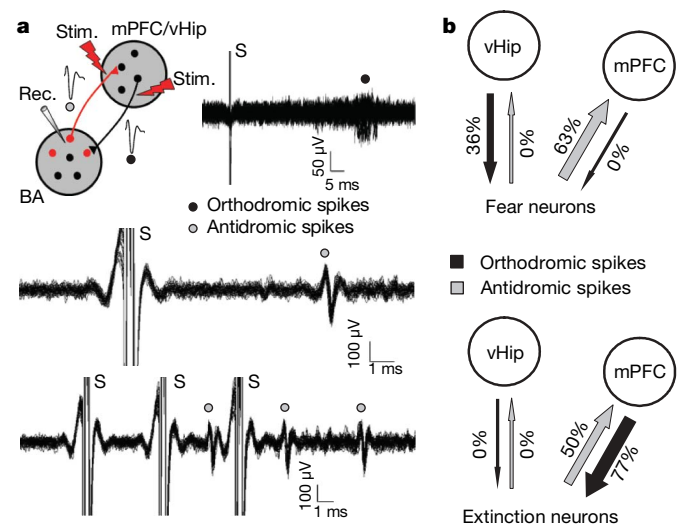
We next addressed the question of whether fear neurons and extinction neurons are anatomically segregated. Comparing the location of electrolytic lesions made by the electrodes from which fear and extinction neurons were recorded did not provide any evidence for anatomical segregation (Supplementary Fig. 4). As a complementary approach, we compared the anatomical distribution of BA neurons activated during exposure to an extinguished or to a non-extinguished CS using the immediate early gene product Fos as an activity-marker. Given the similar numbers of extinction and fear neurons, one would

predict that an extinguished and a non-extinguished CS should induce Fos expression in an equal number of BA neurons with an overlapping anatomical distribution. Consistent with this, we found no difference in the density and anatomical distribution of Fos-positive neurons in animals exposed to an extinguished and a non-extinguished CS (Supplementary Fig. 4). Together, these results suggest that BA fear and extinction neurons are intermingled in a salt-and-pepper-like manner.

Converging evidence supports a role for the mPFC in the consolidation of extinction memory<sup>8,16,17,31</sup>, and for the hippocampus in processing contextual information relevant for the expression and extinction of conditioned fear behaviour<sup>19</sup>. This raises the question as to how fear and extinction neurons in the BA communicate with the mPFC and the hippocampus during context-dependent behavioural transitions. We first addressed the possibility that fear neurons might be excitatory projection neurons whereas extinction neurons might be inhibitory interneurons. However, both fear and extinction neurons exhibited low spontaneous firing rates characteristic of BLA projection neurons<sup>32</sup> (Supplementary Table 1). Consistent with this, analysis of cross-correlations between identified fear or extinction neurons and neighbouring BA neurons revealed short-latency excitatory interactions (Supplementary Fig. 5). To examine whether identified fear and extinction neurons project to, or receive input from, the mPFC and/or the hippocampus, we tested for antidromic activation of BA efferents and orthodromic activation of afferents by using extracellular stimulation electrodes in re-anaesthetized mice (Fig. 5a;



**Figure 4 | Context-dependent fear renewal induces rapid reversal of neuronal activity patterns.** **a**, Experimental protocol. **b**, Summary of behavioural data. Seven days after extinction, extinction memory was tested in the same context in which extinction training took place ( $n = 15$  animals). Both CS<sup>+</sup> and CS<sup>-</sup> elicited low freezing behaviour. Subsequently, mice were placed back into the context in which fear conditioning took place. In this context, exposure to the CS<sup>+</sup> evoked significantly more freezing than CS<sup>-</sup> stimulation. Error bars indicate mean  $\pm$  s.e.m. **c**, Context-dependent changes in CS<sup>+</sup>-evoked firing of fear neurons ( $n = 19$  neurons from 9 mice). Fear neurons exhibited a context-dependent increase in CS<sup>+</sup>-evoked firing in the fear-conditioning context where freezing levels were high ( $P < 0.05$  versus extinction context and versus CS<sup>-</sup>). **d**, Extinction neurons ( $n = 14$  neurons, 8 mice) showed the opposite pattern. Whereas CS<sup>+</sup>-exposure elicited strong firing in the extinction context ( $P < 0.05$  versus fear-conditioning context and versus CS<sup>-</sup>), extinction neurons did not show any CS<sup>+</sup>-evoked responses in the fear-conditioning context. \*\* $P < 0.01$ .



**Figure 5 | Fear neurons and extinction neurons are part of distinct neuronal circuits.** **a**, Use of extracellular stimulation in anaesthetized mice to identify orthodromic and antidromic connections between BA neurons and the mPFC or the hippocampus. Top left: schematic illustrating the placement of stimulating and recording electrodes. Rec., recording electrode; Stim., stimulation electrode. Top right: orthodromic spikes elicited in a BA fear neuron on stimulation of the ventral hippocampus. Orthodromic spikes exhibited a large temporal jitter and high failure rates. Middle: antidromic spikes recorded from a BA extinction neuron in response to mPFC stimulation. Antidromic spikes exhibited low temporal jitter, and followed high frequency (200 Hz) stimulation (bottom). **b**, Top: fear neurons project to the mPFC (5 out of 8 stimulated neurons) and receive input from the hippocampus (5 out of 14 stimulated neurons). vHip, ventral hippocampus. No antidromic responses from the hippocampus (0 out of 14 stimulated neurons) or orthodromic responses from the mPFC (0 out of 8 stimulated neurons) were observed. The graph depicts the percentage of all stimulation experiments in which a particular response was observed in identified fear neurons. Bottom: extinction neurons are reciprocally connected with the mPFC (antidromic responses, 3 out of 6 stimulated neurons; orthodromic responses, 7 out of 9 stimulated neurons,  $P < 0.001$  versus fear neurons). No connections with the hippocampus were observed (0 out of 9 stimulated neurons,  $P < 0.05$  versus fear neurons).

see Methods). These experiments revealed that fear neurons received input from the hippocampus, whereas no connections with the hippocampus were found for extinction neurons ( $P < 0.05$  versus fear neurons; Fig. 5b). Although these findings cannot exclude that some extinction neurons might be contacted by hippocampal afferents, they demonstrate that the probability of receiving hippocampal input is significantly different for fear and extinction neurons. Likewise, fear and extinction neurons were differentially connected with the mPFC. Whereas extinction neurons were reciprocally connected, fear neurons projected to the mPFC, but we did not find any inputs ( $P < 0.001$  versus extinction neurons; Fig. 5b). Extinction-resistant neurons were reciprocally connected to both the mPFC and to the hippocampus (Supplementary Fig. 6). Taken together, these findings indicate that fear and extinction neurons, although co-localized within the same nucleus, not only are functionally specialized but also form part of discrete neuronal circuits.

### BA inactivation prevents behavioural transitions

The observed changes in CS<sup>+</sup>-evoked spike firing of fear and extinction neurons during the extinction and context-dependent renewal of conditioned fear responses could be necessary for the acquisition, storage and/or behavioural expression of the learned information. To distinguish between these possibilities, we used micro-iontophoresis of a fluorescently labelled GABA<sub>A</sub> ( $\gamma$ -aminobutyric acid subtype A) receptor agonist (muscimol) to reversibly inactivate neuronal activity in the BA in a targeted and controlled manner (Fig. 6a). Simultaneous iontophoresis and multi-unit recording revealed that muscimol application silenced neuronal activity in the BA for more than 60 min (Fig. 6b). We first tested whether BA activity was necessary for the acquisition of extinction. Inactivation of the BA completely prevented the decrease in freezing behaviour normally observed during extinction training (Fig. 6c), with no effect on pre-CS freezing levels (not shown). Twenty-four hours later, after wash-out of muscimol, the same animals initially exhibited high freezing levels followed by normal fear extinction, demonstrating that BA inactivation did not merely interfere with the behavioural expression of extinction, nor irreversibly damage BA function (Fig. 6d). These results demonstrate that BA activity is necessary for the acquisition of extinction.

Next, we tested whether BA activity was necessary for the context-dependent renewal of previously extinguished fear responses. Mice exhibiting low freezing levels during recall of extinction memory one week after extinction training were injected with muscimol before renewal. In contrast to control animals injected with the fluorescent label only, muscimol-injected animals exhibited no increase in freezing levels when exposed to the CS<sup>+</sup> in the fear-conditioning context (Fig. 6e). These results demonstrate that BA activity is necessary for context-dependent fear renewal.

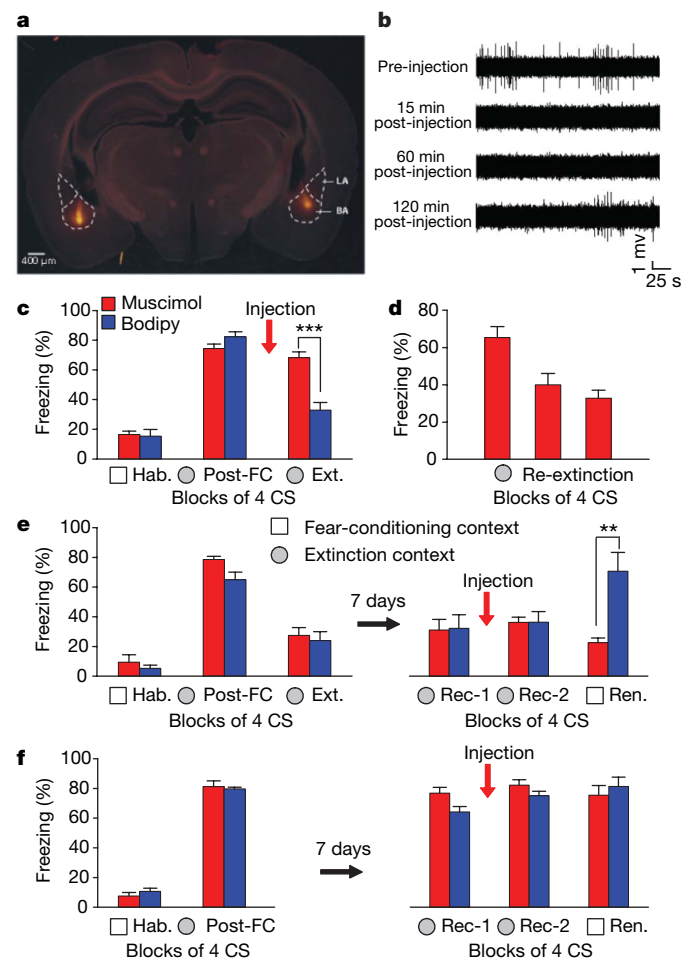
Because muscimol unselectively silences all neurons in the targeted region, the high fear level observed in muscimol-injected mice during extinction learning cannot be accounted for by activity of fear neurons. Conversely, the low fear level displayed by muscimol-injected mice during context-dependent fear renewal cannot be dependent on the activation of extinction neurons. Thus, whereas animals with inactivated BA are able to express high and low fear states, possibly by activation of other parts of the amygdala and the mPFC, they exhibit emotional perseveration (that is, they remained in the emotional state they were in before BA inactivation). This suggests that the BA is unlikely to be associated with the storage, retrieval or expression of conditioned fear and extinction memories, but may instead mediate context-dependent behavioural transitions between low and high fear states.

Thus, silencing of BA activity should have no effect on the retrieval and expression of conditioned and extinguished fear memories when there is no need to change fear levels in a context-dependent manner. Consistent with this scenario, BA inactivation had no effect on the retrieval or expression of consolidated extinction memories (Fig. 6e).

Moreover, in animals that had been fear conditioned one week before, but that did not receive extinction training, muscimol had no effect on the retrieval and expression of the fear memory independently of the context in which they were tested (Fig. 6f).

### Discussion

Our data show that the BA contains distinct populations of neurons for which activity is oppositely correlated with high and low fear behaviour—two converse behavioural states. Although fear and extinction neurons represent relatively small sub-populations within the BA, a rapid switch in the balance of their activity is essential for



**Figure 6 | Targeted inactivation of the BA prevents behavioural changes without affecting memory.** **a**, Epifluorescent image illustrating bilateral targeting of the BA with muscimol covalently attached to a fluorescent tag (bodipy, dipyrromethene boron difluoride). **b**, Simultaneous multi-unit recordings revealed silencing of neuronal activity for up to two hours after muscimol iontophoresis. **c**, Inactivation of the BA before extinction training prevented the acquisition of extinction. Control mice injected with fluorophore only ( $n = 5$ ; blue bars) exhibited significant reduction of freezing levels after extinction training. Muscimol-injected animals ( $n = 11$ ; red bars) showed high freezing levels after extinction. **d**, Twenty-four hours later, in the absence of muscimol, the same animals showed normal acquisition of extinction ( $P < 0.05$ ). Bars illustrate the progressive decrease in freezing behaviour during extinction learning. **e**, Inactivation of the BA prevented context-dependent renewal. Control mice injected with fluorophore only ( $n = 5$ ) exhibited a significant increase in freezing levels on change of context ( $P < 0.05$ ). Muscimol-injected animals ( $n = 5$ ) did not show any context-dependent fear renewal ( $P < 0.01$  versus control). **f**, In the absence of extinction training, BA inactivation did not affect fear memory retrieval. Fluorophore-injected mice ( $n = 4$ ) and muscimol-injected mice ( $n = 5$ ) exhibited equal freezing levels during CS<sup>+</sup> exposure in the fear-conditioning context one week after fear conditioning. Error bars indicate mean  $\pm$  s.e.m. \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

triggering behavioural transitions during extinction and context-dependent fear renewal. Although intermingled within the BA, fear and extinction neurons are differentially connected with the hippocampus and the mPFC, two brain areas previously implicated in extinction and context-dependent renewal of conditioned fear responses. In keeping with the proposed role of the ventral hippocampus in mediating context-dependent renewal of fear behaviour in animals subjected to extinction<sup>23</sup>, we found that hippocampal input to the BA selectively targets fear neurons over extinction neurons. Thus, hippocampal input to BA fear neurons may override the retrieval of extinction memory allowing for fear expression after a particular CS has undergone extinction. Extinction neurons, in turn, are bi-directionally connected with the mPFC and are switched on during extinction training. This indicates that they may be upstream of a previously identified population of mPFC neurons thought to mediate consolidation of extinction memory, because they are activated by an extinguished CS during recall, but not during the acquisition of extinction<sup>8</sup>.

Previous findings demonstrate that the BLA is not critical for triggering behavioural transitions during reversal learning in a two-odour-discrimination task<sup>33,34</sup>. Nevertheless, abnormally persistent BLA activity induced by orbitofrontal cortex lesions<sup>33</sup> or repeated cocaine administration<sup>34</sup> interferes with reversal learning. This suggests that, whereas the BLA can only veto slow behavioural transitions during more complex reversal learning tasks, it is actively involved in situations requiring rapid context-dependent switching between two converse behavioural states.

How might activity of BA fear and extinction neurons mediate behavioural transitions? In keeping with a role for the amygdala in facilitating network function and memory formation in other parts of the brain<sup>35–37</sup>, a possible interpretation is that BA fear and extinction neurons might drive or facilitate the induction of synaptic plasticity in their respective target areas. Moreover, whereas previous studies using pre-fear-conditioning lesions came to the conclusion that the BA does not contribute to the acquisition or the expression of conditioned fear<sup>38–41</sup> (but see ref. 42), a recent analysis using post-fear-conditioning lesions indicates that the BA also contributes to the consolidation of long-term fear memories<sup>41</sup>. This suggests that repeated activity of BA fear neurons, over longer-time periods, may be required for fear memory consolidation.

Our findings are consistent with the idea that in mammals, as in invertebrates<sup>43,44</sup>, switches between appropriate behavioural states can be driven by discrete neuronal circuits. It may be a general principle of the functional micro-architecture of the nervous system in diverse species, that circuits mediating switches between distinct behavioural states are located in close anatomical proximity thereby allowing for local interactions. However, it remains to be shown how fear and extinction neurons interact locally. Finally, our results also suggest that context-dependent recovery of extinguished fear behaviour in humans<sup>45</sup>, which represents a major clinical obstacle for the therapy of certain anxiety disorders<sup>46</sup>, might be modulated by tipping the balance of activity between specific neuronal circuits.

## METHODS SUMMARY

**Behaviour and pharmacological inactivations.** Mice were submitted to a discriminative auditory fear-conditioning paradigm in which the CS<sup>+</sup>, but not the CS<sup>−</sup>, was paired with a US (mild foot shock). Extinction training was performed over two days in a different context<sup>14</sup>. One week later, mice were placed in the extinction context for recall of extinction, and in the original conditioning context for fear renewal. Freezing behaviour was quantified during each behavioural session using an automatic infrared beam detection system as described previously<sup>47</sup>. Bilateral inactivation of the BA was achieved using micro-iontophoretic injection of fluorescently labelled muscimol before extinction training or context-dependent fear renewal.

**Electrophysiological recordings and analysis.** Individual neurons were recorded extracellularly in freely behaving mice during fear conditioning, extinction, recall of extinction and context-dependent fear renewal. Spikes of individual neurons were sorted by time–amplitude window discrimination and

template matching as described previously<sup>47,48</sup>. Cluster quality and unit stability were verified by quantifying the cluster separation and the stability of the average waveform shape over time<sup>48,49</sup> (Supplementary Fig. 7). Unit isolation was verified using auto- and cross-correlation histograms. Spike rasters and histograms were constructed by aligning sweeps relative to the CS onset, and CS-evoked responses were normalized to baseline activity using a z-score transformation. Antidromic and orthodromic spikes evoked by extracellular stimulations of the mPFC or the ventral hippocampus were recorded in previously identified neurons in anaesthetized mice.

**Immunohistochemistry.** Mice were submitted to an auditory fear-conditioning paradigm in which only one CS was used. Two hours after the end of training, mice exposed to the CS only, extinguished and non-extinguished mice were transcardially perfused, and their brains collected for immunohistochemical staining of the immediate early gene product Fos as previously described<sup>26</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 19 March; accepted 11 June 2008.**

**Published online 9 July 2008.**

- Quirk, G. J., Repa, C. & LeDoux, J. E. Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. *Neuron* **15**, 1029–1039 (1995).
- Collins, D. R. & Paré, D. Differential fear conditioning induces reciprocal changes in the sensory responses of lateral amygdala neurons to the CS(+) and CS(−). *Learn. Mem.* **7**, 97–103 (2000).
- Rosenkranz, J. A. & Grace, A. A. Dopamine-mediated modulation of odour-evoked amygdala potentials during Pavlovian conditioning. *Nature* **417**, 282–287 (2002).
- Goossens, K. A., Hobin, J. A. & Maren, S. Auditory-evoked spike firing in the lateral amygdala and Pavlovian fear conditioning: mnemonic code or fear bias? *Neuron* **40**, 1013–1022 (2003).
- LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
- Maren, S. & Quirk, G. J. Neuronal signaling of fear memory. *Nature Rev. Neurosci.* **5**, 844–852 (2004).
- Quirk, G. J., Russo, G. K., Barron, J. L. & Lebron, K. The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *J. Neurosci.* **20**, 6225–6231 (2000).
- Milad, M. R. & Quirk, G. J. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* **420**, 70–74 (2002).
- Rosenkranz, J. A., Moore, H. & Grace, A. A. The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. *J. Neurosci.* **23**, 11054–11064 (2003).
- Berretta, S., Pantazopoulos, H., Caldera, M., Pantazopoulos, P. & Paré, D. Infralimbic cortex activation increases c-Fos expression in intercalated neurons of the amygdala. *Neuroscience* **132**, 943–953 (2005).
- Paré, D., Quirk, G. J. & LeDoux, J. E. New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.* **92**, 1–9 (2004).
- Falls, W. A., Miserendino, M. J. & Davis, M. Extinction of fear-potentiated startle: blockade by infusion of an NMDA antagonist into the amygdala. *J. Neurosci.* **12**, 854–863 (1992).
- Lu, K. T., Walker, D. L. & Davis, M. Mitogen-activated protein kinase cascade in the basolateral nucleus of amygdala is involved in extinction of fear-potentiated startle. *J. Neurosci.* **21**, RC162 (2001).
- Herry, C., Trillieff, P., Micheau, J., Lüthi, A. & Mons, N. Extinction of auditory fear conditioning requires MAPK/ERK activation in the basolateral amygdala. *Eur. J. Neurosci.* **24**, 261–269 (2006).
- Sotres-Bayon, F., Bush, D. E. & LeDoux, J. E. Acquisition of fear extinction requires activation of NR2B-containing NMDA receptors in the lateral amygdala. *Neuropsychopharmacology* **32**, 1929–1940 (2007).
- Quirk, G. J., Garcia, R. & Gonzalez-Lima, F. Prefrontal mechanisms in extinction of conditioned fear. *Biol. Psychiatry* **60**, 337–343 (2006).
- Myers, K. M. & Davis, M. Mechanisms of fear extinction. *Mol. Psychiatry* **12**, 120–150 (2007).
- Bouton, M. E. Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biol. Psychiatry* **52**, 976–986 (2002).
- Bouton, M. E., Westbrook, R. F., Corcoran, K. A. & Maren, S. Contextual and temporal modulation of extinction: Behavioral and biological mechanisms. *Biol. Psychiatry* **60**, 352–360 (2006).
- Pitkänen, A., Pikkarainen, M., Nurminen, N. & Ylinen, A. Reciprocal connections between the amygdala and the hippocampal formation, perirhinal cortex, and postnatal cortex in rat. A review. *Ann. NY Acad. Sci.* **911**, 369–391 (2000).
- Corcoran, K. A. & Maren, S. Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. *J. Neurosci.* **21**, 1720–1726 (2001).
- Corcoran, K. A., Desmond, T. J., Frey, K. A. & Maren, S. Hippocampal inactivation disrupts the acquisition and contextual encoding of fear extinction. *J. Neurosci.* **25**, 8978–8987 (2005).
- Hobin, J. A., Ji, J. & Maren, S. Ventral hippocampal muscimol disrupts context-specific fear memory retrieval after extinction in rats. *Hippocampus* **16**, 174–182 (2006).



24. McDonald, A. J. Organization of amygdaloid projections to the prefrontal cortex and associated striatum in the rat. *Neuroscience* **44**, 1–14 (1991).
25. McDonald, A. J., Mascagni, F. & Guo, L. Projections of the medial and lateral prefrontal cortices to the amygdala: a *Phaseolus vulgaris* leucoagglutinin study in the rat. *Neuroscience* **71**, 55–75 (1996).
26. Herry, C. & Mons, N. Resistance to extinction is associated with impaired immediate early gene induction in medial prefrontal cortex and amygdala. *Eur. J. Neurosci.* **20**, 781–790 (2004).
27. Muramoto, K., Ono, T., Nishijo, H. & Fukuda, M. Rat amygdaloid neuron responses during auditory discrimination. *Neuroscience* **52**, 621–636 (1993).
28. Maren, S., Poremba, A. & Gabriel, M. Basolateral amygdaloid multi-unit neuronal correlates of discriminative avoidance learning in rabbits. *Brain Res.* **549**, 311–316 (1991).
29. Repa, J. C. *et al.* Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nature Neurosci.* **4**, 724–731 (2001).
30. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: Implications of a quantitative analysis. *Proc. Natl Acad. Sci. USA* **101**, 13124–13131 (2004).
31. Herry, C. & Garcia, R. Prefrontal cortex long-term potentiation, but not long-term depression, is associated with the maintenance of extinction of learned fear in mice. *J. Neurosci.* **22**, 577–583 (2002).
32. Likhtik, E., Pelletier, J. G., Popescu, A. T. & Paré, D. Identification of basolateral amygdala projection cells and interneurons using extracellular recordings. *J. Neurophysiol.* **96**, 3257–3265 (2006).
33. Stalnaker, T. A., Franz, T. M., Singh, T. & Schoenbaum, G. Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron* **54**, 51–58 (2007).
34. Stalnaker, T. A. *et al.* Cocaine-induced decision-making deficits are mediated by misrouting in basolateral amygdala. *Nature Neurosci.* **10**, 949–951 (2007).
35. Paré, D. Role of the basolateral amygdala in memory consolidation. *Prog. Neurobiol.* **70**, 409–420 (2003).
36. McGaugh, J. L. The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci.* **27**, 1–28 (2004).
37. Paz, R., Pelletier, J. G., Bauer, E. P. & Paré, D. Emotional enhancement of memory via amygdala-driven facilitation of rhinal interactions. *Nature Neurosci.* **9**, 1321–1329 (2006).
38. Amorapanth, P., LeDoux, J. E. & Nader, K. Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nature Neurosci.* **3**, 74–79 (2000).
39. Nader, K., Majidishad, P., Amorapanth, P. & LeDoux, J. E. Damage to the lateral and central, but not other, amygdaloid nuclei prevents the acquisition of auditory fear conditioning. *Learn. Mem.* **8**, 156–163 (2001).
40. Sotres-Bayon, F., Bush, D. E. & LeDoux, J. E. Emotional perseveration: an update on prefrontal-amygdala interactions in fear extinction. *Learn. Mem.* **11**, 525–535 (2004).
41. Anglada-Figueroa, D. & Quirk, G. J. Lesions of the basal amygdala block expression of conditioned fear but not extinction. *J. Neurosci.* **25**, 9680–9685 (2005).
42. Goossens, K. A. & Maren, S. Contextual and auditory fear conditioning are mediated by the lateral, basal, and central amygdaloid nuclei in rats. *Learn. Mem.* **8**, 148–155 (2001).
43. Jing, J. & Gillette, R. Escape swim network interneurons have diverse roles in behavioral switching and putative arousal in *Pleurobranchanea*. *J. Neurophysiol.* **83**, 1346–1355 (2000).
44. Yapici, N., Kim, Y.-J., Ribeiro, C. & Dickson, B. J. A receptor that mediates the post-mating switch in *Drosophila* reproductive behavior. *Nature* **451**, 33–37 (2007).
45. Milad, M. R., Orr, S. P., Pitman, R. K. & Rauch, S. L. Context modulation of memory for fear extinction in humans. *Psychophysiology* **42**, 456–464 (2005).
46. Rodriguez, B. I., Craske, M. G., Mineka, S. & Hladek, D. Context-specificity of relapse: effects of therapist and environmental context on return of fear. *Behav. Res. Ther.* **37**, 845–862 (1999).
47. Herry, C. *et al.* Processing of temporal unpredictability in human and animal amygdala. *J. Neurosci.* **27**, 5958–5966 (2007).
48. Nicolelis, M. A. L. *et al.* Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc. Natl Acad. Sci. USA* **100**, 11041–11046 (2003).
49. Jackson, A. & Fetz, E. E. Compact movable microwire array for long-term chronic unit recordings in cerebral cortex of primates. *J. Neurophysiol.* **98**, 3109–3118 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Friedrich, A. Matus and all members of the Lüthi laboratory for discussions and critical comments on the manuscript and E. Oakeley and M. Stadler for advice on statistical analysis and programming. This work was supported by the Austrian Science Fund (FWF), the Novartis Institutes for Biomedical Research, and the Novartis Research Foundation.

**Author Contributions** C.H., S.C., V.S., L.D. and C.M. performed the experiments and analysed the data. C.H., S.C. and A.L. designed the experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.L. ([andreas.luthi@fmi.ch](mailto:andreas.luthi@fmi.ch)) or C.H. ([cyril.herry@fmi.ch](mailto:cyril.herry@fmi.ch)).

## METHODS

**Animals.** Male C57BL/6J mice (3 months old; RCC Ltd) were individually housed for 7 days before all experiments, under a 12 h light/dark cycle, and provided with food and water *ad libitum*. All animal procedures were performed in accordance with institutional guidelines and were approved by the Veterinary Department of the Canton of Basel-Stadt.

**Behaviour.** Fear conditioning and extinction took place in two different contexts (context A and B). The conditioning and extinction boxes and the floor were cleaned with 70% ethanol or 1% acetic acid before and after each session, respectively. To score freezing behaviour, an automatic infrared beam detection system placed on the bottom of the experimental chambers (Coulbourn Instruments) was used. The animals were considered to be freezing if no movement was detected for 2 s. On day 1, mice were submitted to a habituation session in context A, in which they received 4 presentations of the CS<sup>+</sup> and the CS<sup>-</sup> (total CS duration of 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2-ms rise and fall; pip frequency: 7.5 kHz or white-noise, 80 dB sound pressure level). Discriminative fear conditioning was performed on the same day by pairing the CS<sup>+</sup> with a US (1-s foot shock, 0.6 mA, 5 CS<sup>+</sup>/US pairings; inter-trial interval: 20–180 s). The onset of the US coincided with the offset of the CS<sup>+</sup>. The CS<sup>-</sup> was presented after each CS<sup>+</sup>/US association but was never reinforced (5 CS<sup>-</sup> presentations, inter-trial interval: 20–180 s). The frequencies used for CS<sup>+</sup> and CS<sup>-</sup> were counterbalanced across animals. On day 2 and day 3, conditioned mice were submitted to extinction training in context B, during which they received 4 and 12 presentations of the CS<sup>-</sup> and the CS<sup>+</sup>, respectively. Recall of extinction and context-dependent fear renewal were tested 7 days later in context B and A, respectively, with 4 presentations of the CS<sup>-</sup> and the CS<sup>+</sup>. Pharmacological experiments were performed using the same conditioning and extinction protocol except for one group of mice that was not submitted to extinction training but tested for conditioned fear with 4 CS<sup>-</sup> and 4 CS<sup>+</sup> presentations on day 2 in context B. Seven days later, mice were submitted to 2 sessions of extinction recall 5 h apart in context B (4 presentations of each CS for each session). Finally, 10 min after the second recall session, mice were submitted to 4 CS<sup>-</sup> and 4 CS<sup>+</sup> presentations in context A for context-dependent fear renewal.

For discriminative extinction, mice were habituated on day 1 to 4 presentations of two different CS in context A (total CS duration of 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2 ms rise and fall; pip frequency: 7.5 kHz or white-noise, 80 dB sound pressure level). Both CS were subsequently paired with a US (1-s foot shock, 0.6 mA, 5 CS/US pairings for each CS; inter-trial interval: 20–180 s). The onset of the US coincided with the offset of the CS. On days 3 and 4, only one of the two CS was extinguished by 16 and 12 presentations in context B, respectively. At the end of the second extinction session, mice were exposed to 4 presentations of the non-extinguished CS in context B.

**Surgery and recordings.** Mice were anesthetized with isoflurane (induction 5%, maintenance 2.5%) in O<sub>2</sub>. Body temperature was maintained with a heating pad (CMA/150, CMA/Microdialysis). Mice were secured in a stereotaxic frame and unilaterally implanted in the amygdala with a multi-wire electrode aimed at the following coordinates<sup>30</sup>: 1.7 mm posterior to bregma; ±3.1 mm lateral to midline; and 4 mm to 4.3 mm deep from the cortical surface. The electrodes consisted of 8 to 16 individually insulated nichrome wires (13-μm inner diameter, impedance 1–3 MΩ; California Fine Wire) contained in a 26 gauge stainless steel guide canula. The wires were attached to a 10 pin to 18 pin connector (Omnetics). The implant was secured using cyanoacrylate adhesive gel. After surgery mice were allowed to recover for 7 days. Analgesia was applied before, and during the 3 days after, surgery (Metacam, Boehringer). Electrodes were connected to a headstage (Plexon) containing eight to sixteen unity-gain operational amplifiers. The headstage was connected to a 16-channel computer-controlled preamplifier (gain 100×, bandpass filter from 150 Hz to 9 kHz, Plexon). Neuronal activity was digitized at 40 kHz and bandpass filtered from 250 Hz to 8 kHz, and was isolated by time–amplitude window discrimination and template matching using a Multichannel Acquisition Processor system (Plexon). At the conclusion of the experiment, recording sites were marked with electrolytic lesions before perfusion, and electrode locations were reconstructed with standard histological techniques.

**Single-unit spike sorting and analysis.** Single-unit spike sorting was performed using Off-Line Spike Sorter (OFSS, Plexon) as described<sup>47,48</sup> (Supplementary Fig. 7). Principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces, and clusters containing similar valid waveforms were manually defined. A group of waveforms was considered to be generated from a single neuron if it defined a discrete cluster in principal component space that was distinct from clusters for other units and if it displayed a clear refractory period (>1 ms) in the auto-correlogram histograms. In addition, two parameters were used to quantify the overall separation between identified clusters in a particular channel. These parameters include the J3 statistic,

which corresponds to the ratio of between-cluster to within-cluster scatter, and the Davies–Bouldin validity index (DB), which reflects the ratio of the sum of within-cluster scatter to between-cluster separation<sup>48</sup>. High values for the J3 and low values for the DB are indicative of good single-unit isolation (Supplementary Fig. 7). Control values for this statistics were obtained by artificially defining two clusters from the centred cloud of points in the principal component space from channels in which no units could be detected. Template waveforms were then calculated for well-separated clusters and stored for further analysis. Clusters of identified neurons were analysed offline for each recording session using principal component analysis and a template-matching algorithm. Only stable clusters of single units recorded over the time course of the entire behavioural training were considered. Long-term single-unit stability isolation was first evaluated using Wavetracker (Plexon) in which principal component space-cylinders were calculated from a 5 min segment of data spontaneously recorded before any training session. Straight cylinders suggest that the same set of single units was recorded during the entire training session (Supplementary Fig. 7).

Second, we quantitatively evaluated the similarity of waveform shape by calculating linear correlation (*r*) values between average waveforms obtained over training days<sup>49</sup> (Supplementary Fig. 7). As a control, we computed the *r* values from average waveforms of different neurons.

Third, for each unit we used correlation analysis to quantitatively compare the similarity of waveform shape during CS<sup>+</sup>-stimulation and during a 60 s period of spontaneous activity recorded before each behavioural session (Supplementary Fig. 8). To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at a time that the reference neuron fires, only one of the two neurons was considered for further analysis. CS-induced neural activity was calculated by comparing the firing rate after stimulus onset with the firing rate recorded during the 500 ms before stimulus onset (bin size, 20 ms; averaged over blocks of 4 CS presentations consisting of 108 individual sound pips in total) using a *z*-score transformation. *z*-score values were calculated by subtracting the average baseline firing rate established over the 500 ms preceding stimulus onset from individual raw values and by dividing the difference by the baseline standard deviation. Only CS-excited neurons were considered for analysis. Classification of units was performed by comparing the largest significant *z*-score values within 100 ms after CS-onset during post-fear conditioning and extinction sessions according to the freezing levels. For high-fear states, the entire post-fear conditioning session was analysed, whereas, for low-fear states, analysis was restricted to the block of 4 CS presentations during which the fear level was the lowest. A unit was classified as a fear neuron if it exhibited a significant *z*-score value after fear conditioning (when freezing levels were high), but no significant *z*-score value after extinction (when freezing levels were low), and vice versa for extinction neurons. Finally, units were classified as extinction-resistant neurons if they displayed a significant *z*-score value during both post-fear conditioning and extinction sessions, independently of freezing levels. For statistical analysis, *z*-score comparisons were performed using the average *z*-score value calculated during the 40 ms after CS-onset. In cases in which shorter or longer CS-evoked activity was observed, the average *z*-score was calculated during the 20 ms and 80 ms after CS-onset, respectively. To identify the trial in which individual neurons changed their CS-evoked responses during fear conditioning and extinction, we applied a change point analysis algorithm<sup>30</sup>. Change point analysis identifies the trial(s) exhibiting a significant change in neuronal activity or freezing behaviour relative to the preceding trials. Change points are graphically represented by a change in the slope of a plot showing the cumulative sums of the averaged and normalized *z*-score and freezing values. Statistical analyses were performed using paired Student's *t*-tests *post hoc* comparisons at the *P* < 0.05 level of significance unless indicated otherwise. Results are presented as mean ± s.e.m.

**Extracellular stimulation.** To determine the connectivity of recorded neurons, we used extracellular stimulation of the mPFC and the vHip in a subset of animals. At the end of the training procedure, animals were anesthetized using urethane (1.4 g kg<sup>-1</sup>), and concentric stimulating electrodes (FHC) were lowered in the mPFC (2 mm anterior to bregma; ±0.3 mm lateral to midline; and 1.6 mm to 2 mm deep from the cortical surface) and the ventral hippocampus (3.6 mm posterior to bregma; ±3.1 mm lateral to midline; and 4 mm to 4.2 mm deep from the cortical surface). During the experiments, the stimulation electrodes were advanced in steps of 5 μm by a motorized micromanipulator (David Kopf Instruments), and BA-evoked responses were recorded. Stimulation-induced and spontaneous spikes were sorted using principal component analysis and template matching. The similarity of stimulation-induced spike waveforms was quantitatively compared to the waveforms of units previously identified in the awake animal and recorded on the same wire using correlation analysis (Supplementary Fig. 9). To be classified as antidromic, evoked responses had to meet at least two out of three criteria: (1) stable latency (<0.3 ms jitter), (2)

collision with spontaneously occurring spikes, and (3) the ability to follow high-frequency stimulation (200 Hz). At the end of the experiments, stimulating sites were marked with electrolytic lesions before perfusion, and electrode locations were reconstructed with standard histological techniques. For each stimulation site, orthodromic and antidromic response probabilities of fear and extinction neurons were analysed using binomial statistics, with  $P < 0.05$  indicating non-random connectivity.

**Muscimol iontophoresis.** Muscimol micro-iontophoresis injection was performed in chronically implanted animals. Single-barrel micropipettes with a tip diameter of 10 to 15  $\mu\text{m}$  were cut at 1 cm length and filled with a solution containing muscimol covalently coupled to a fluorophore (Muscimol-Bodipy-TMR conjugated, Invitrogen; 5 mM in phosphate buffered saline (PBS) 0.1 M, DMSO 40%) or with bodipy alone (Invitrogen; 5 mM in PBS 0.1 M, DMSO 40%). Mice were bilaterally implanted at the following coordinates (according to ref. 50): 1.7 mm posterior to bregma;  $\pm 3.1$  mm lateral to midline; and 4 mm to 4.3 mm deep from the cortical surface. Chlorided silver wires were inserted in each micropipette and attached to a connector. A third silver wire screwed onto the skull and attached to the connector served as a reference electrode. The entire miniaturized assembly was secured using cyanoacrylate adhesive gel. After surgery, mice were allowed to recover for 2 days. On the injection day, iontophoretic applications were performed by means of cationic current ( $+12 \mu\text{A}$  to  $+15 \mu\text{A}$ ) for 15 min per side using a precision current source device (Stoelting). Mice were submitted to the behavioural procedure 5 min after the end of iontophoretic injections and were immediately perfused at the end of the experiments. Brains were collected for further histological analysis. Serial slices containing the amygdala were imaged at  $\times 5$  using an epifluorescence stereo microscope (Leica), and the location and the extent of the injections were controlled. Mice were included in the analysis only if they presented a bilateral injection targeting exclusively the BA and if the targeted injections cover at least 25% of the BA. Statistical analyses were performed using paired and unpaired Student's *t*-tests *post hoc* comparisons at the  $P < 0.05$  level of significance. Results are presented as mean  $\pm$  s.e.m.

**Immunohistochemistry.** Mice were transcardially perfused with ice-cold 4% paraformaldehyde in 0.1 M PBS 120 min after the onset of the training session<sup>26</sup>. Brains were prepared for immunohistochemistry using primary polyclonal rabbit anti-c-Fos antibody (Calbiochem; anti-c-Fos, Ab-5, 4-17, rabbit pAb, PC38; 1:20,000 dilution). A fluorescent-dye-coupled goat anti-rabbit antibody (Invitrogen; Alexa-Fluor 633; 1:1,000 in PBS) was used as secondary antibody. Stained slices were imaged at  $\times 40$  using an LSM 510 Meta confocal microscope (Carl Zeiss Inc.). Quantitative analysis of c-Fos-positive nuclei was performed using a computerized image analysis system (Imaris 4.2, Bitplane). Structures were defined according to ref. 50. Immunoreactive neurons were counted bilaterally using a minimum of three sections per hemisphere per animal. Statistical analyses were performed using unpaired Student's *t*-tests at the  $P < 0.05$  level of significance. Results are presented as mean  $\pm$  s.e.m.

50. Franklin, K. J. B. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* (Academic, 1997).



# The identification of liquid ethane in Titan's Ontario Lacus

R. H. Brown<sup>1</sup>, L. A. Soderblom<sup>2</sup>, J. M. Soderblom<sup>1</sup>, R. N. Clark<sup>3</sup>, R. Jaumann<sup>4</sup>, J. W. Barnes<sup>5</sup>, C. Sotin<sup>6</sup>, B. Buratti<sup>6</sup>, K. H. Baines<sup>6</sup> & P. D. Nicholson<sup>7</sup>

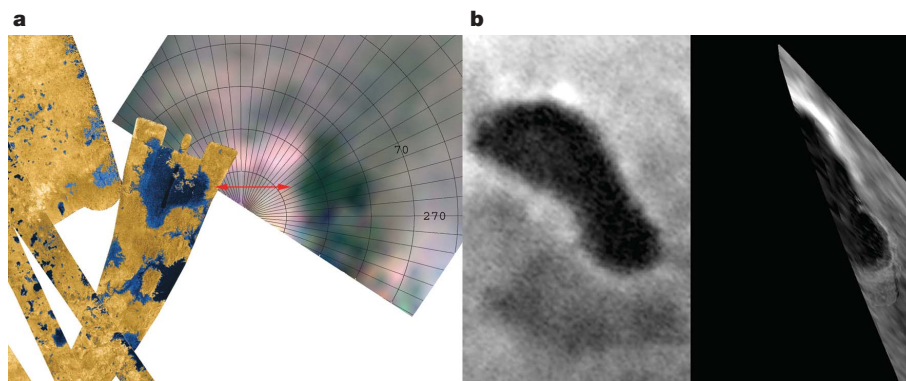
Titan was once thought to have global oceans of light hydrocarbons on its surface<sup>1–5</sup>, but after 40 close flybys of Titan by the Cassini spacecraft, it has become clear that no such oceans exist<sup>6</sup>. There are, however, features similar to terrestrial lakes and seas<sup>7</sup>, and widespread evidence for fluvial erosion<sup>8,9</sup>, presumably driven by precipitation of liquid methane from Titan's dense, nitrogen-dominated atmosphere<sup>10</sup>. Here we report infrared spectroscopic data, obtained by the Visual and Infrared Mapping Spectrometer<sup>11</sup> (VIMS) on board the Cassini spacecraft, that strongly indicate that ethane, probably in liquid solution with methane, nitrogen and other low-molecular-mass hydrocarbons, is contained within Titan's Ontario Lacus.

Saturn's moon Titan has a thick atmosphere that makes studies of its surface difficult, except through specific atmospheric windows<sup>12,13</sup>, as a result of strong atmospheric absorptions and a ubiquitous haze<sup>10,14,15</sup>. Imaging spectroscopy is used to determine the chemical composition of the surfaces and atmospheres of planetary bodies<sup>16</sup>, and the VIMS<sup>11</sup> instrument on board the Cassini spacecraft is capable of such measurements. VIMS was used during Cassini's 38th close flyby of Titan (T38) to observe Ontario Lacus, a lake-like

feature in Titan's south-polar region found with the Cassini Imaging Science Subsystem (ISS) narrow-angle camera in mid-2005. During Cassini's T25 flyby, both VIMS and Cassini's radar instrument were used to observe a northern lake near 78° N, 250° W (ref. 17).

The spectral data obtained for the northern lake were too noisy for compositional analysis, but they indicate that it has an extremely low reflectance in the near-infrared (Fig. 1), consistent with that of Ontario Lacus. Further compositional studies of this lake and others in the north will have to wait until they are directly illuminated in northern spring during Cassini's extended missions.

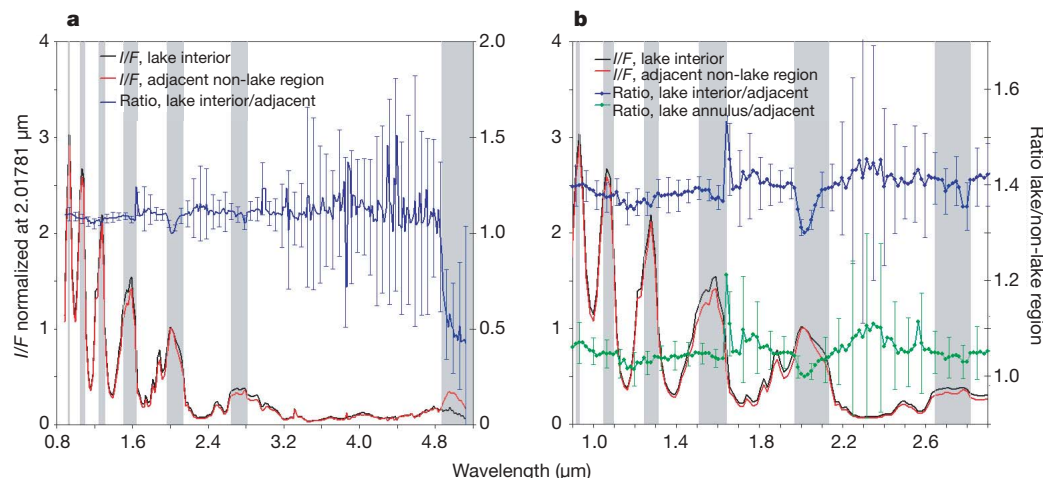
In contrast, the observations of Ontario Lacus with VIMS were of much higher quality, and were obtained in favourable viewing geometry and at much higher spatial resolution. The 'shoreline' of Ontario Lacus (Fig. 1) is quite bright both in the visible and at 5  $\mu$ m relative to the much darker interior of the lake. There is also an annular area (hereafter referred to as the 'beach') just inside the shoreline that is intermediate in brightness and whose inner edge roughly follows the shape of Ontario's shoreline. The reflectivity of this region at 2 and 5  $\mu$ m is intermediate to that of the interior of Ontario Lacus and its shoreline.



**Figure 1 | VIMS and radar images of the lake at 78° N, 250° W, and ISS and VIMS images of Ontario Lacus.** The red arrow indicates the same lake-like feature seen in radar (left side of **a**) and VIMS (right side of **a**). The VIMS image is congruent with the boundaries of the lake as seen in the radar image. The data were obtained at very large illumination angles (the lake was actually 5–10° past the terminator). VIMS observations of the northern lake were possible only because it was illuminated by light scattered past the terminator by Titan's atmosphere. Although the VIMS observations were far from ideal, the northern lake is well defined and is quite clearly the same lake as that imaged at much higher spatial resolution by the Cassini radar<sup>7</sup>. Note that the brightness distribution in the lake's interior as seen in the VIMS image roughly matches the intensity of the radar return as indicated in

shades of blue in the radar image. Colour in the VIMS image was constructed by assigning a band near 1.3  $\mu$ m to the blue channel, a band near 1.6  $\mu$ m to the green channel and a band near 2.0  $\mu$ m to the red channel. The ISS image (left side of **b**) from mid-2005 is compared with the VIMS image (right side of **b**). The VIMS image (taken during the T38 flyby, which occurred on 2007 December 4, at a closest approach altitude of 1,100 km, yielding a maximum resolution of about 500 m per pixel) is a co-addition of 11 channels in the 5- $\mu$ m window. The 'beach' is the thin annular area in the lower right of the image in **b**. It is intermediate in albedo between the lake interior and the surrounding shoreline, and is roughly congruent with the dark boundary of the lake interior and with the bright line delineating the shoreline.

<sup>1</sup>Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA. <sup>2</sup>US Geological Survey, Flagstaff, Arizona 86001, USA. <sup>3</sup>US Geological Survey, Denver, Colorado 80225, USA. <sup>4</sup>Deutsches Zentrum für Luft- und Raumfahrt, 12489 Berlin, Germany. <sup>5</sup>NASA Ames Research Center, Moffett Field, California 94035, USA. <sup>6</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91107, USA. <sup>7</sup>Department of Astronomy, Cornell University, Ithaca, New York 14853, USA.

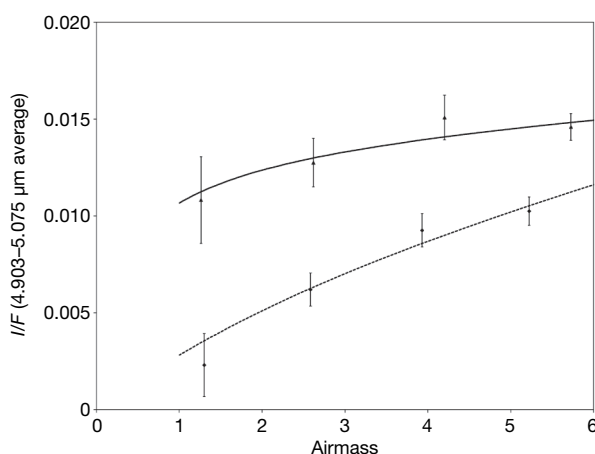


**Figure 2 | Spectra of regions in and around Ontario Lacus.** **a**, The  $I/F$  spectra ( $I$  is the observed specific intensity and  $\pi F$  is the incident solar flux) have been normalized to 1.0 at 2.0178  $\mu\text{m}$ , and the scale is on the left axis. The scale for the ratio spectrum is on the right axis. Even though the lake is intrinsically darker than its surroundings, the ratio is greater than 1.0 because of the normalization. This was done for clarity in the figures. **b**, A similar approach was used and a ratio spectrum for the 'beach' area is

included with a split right axis, again for the sake of clarity. Error bars have been shown for only every third point, again to improve clarity. The effectiveness of the atmospheric band removal is confirmed because the strong methane absorptions are nearly cancelled in the spectral ratio. The vertical grey bands in both panels denote the spectral channels within atmospheric windows through which VIMS can resolve surface detail. Error bars are  $1\sigma$ .

Titan's atmospheric windows at 2.0, 2.7 and 5.0  $\mu\text{m}$  are amenable to surface composition studies<sup>13,18–20</sup>, and within these three windows we find spectral features diagnostic of the composition of the contents of Ontario Lacus. Our  $I/F$  (or radiance factor) spectra of the dark interior of Ontario Lacus (Fig. 2) show strong absorptions that are due mostly to methane gas in Titan's atmosphere, as well as subtle variations within the windows at 2.0, 2.7 and 5.0  $\mu\text{m}$  as a result of photons scattered by Titan's surface. To isolate the spectral differences between Ontario's interior and its shoreline, and to remove atmospheric absorptions, we have constructed a ratio of representative spectra of the Ontario Lacus interior and a nearby area (with nearly identical viewing geometry) outside of the lake.

There are four apparent absorption features in the ratio spectrum (Fig. 2) that appear in only three of seven of Titan's atmospheric windows, indicating that the detailed shapes of the atmospheric windows introduce no artefact absorptions. Thus, the remaining absorptions seen in the ratio are almost certainly associated with Ontario Lacus.



**Figure 3 | Plot of 5- $\mu\text{m}$  brightness against airmass for the lake interior and an adjacent region outside of Ontario Lacus.** The lake interior is denoted with diamonds (lower curve) and the adjacent area by triangles (upper curve). The lines show simple photometric models fitted to the points. Airmass here is the classical astronomical definition of atmospheric path length, commonly called  $\sec(z)$ , where  $z$  is the zenith angle referenced to the surface normal. Error bars are  $1\sigma$ .

In order of increasing wavelength, the first of these bands is centred at 2.018  $\mu\text{m}$ , with a roughly gaussian profile. There is a second feature centred near 2.11  $\mu\text{m}$  that is quite weak and shows mostly as a depression in the long-wavelength wing of the 2.018- $\mu\text{m}$  band. The third band is centred near 2.79  $\mu\text{m}$  and is also roughly gaussian. The fourth feature is a very steep decline in reflectance near 4.8  $\mu\text{m}$ , which continues downwards out to 5.17  $\mu\text{m}$ . Within the region 4.8–5.17  $\mu\text{m}$  there also exist apparent narrow absorption features near 4.9 and 5.05  $\mu\text{m}$  whose significance is not clear.

An important indicator of whether a given absorption band in the Ontario Lacus spectrum pertains to material within the lake is whether the band is spatially correlated with the lake. In the 5- $\mu\text{m}$  image of Ontario Lacus (Fig. 1), the interior of the lake and the 'beach' are congruent to, but clearly have a much lower reflectivity than, the shoreline. Maps of the depths of the bands at 2.018 and 2.79  $\mu\text{m}$  have similar characteristics, although the weaker and noisier 2.79- $\mu\text{m}$  band shows much less correlation with the lake. We are therefore not convinced that the 2.79- $\mu\text{m}$  feature is real and do not discuss it further here.

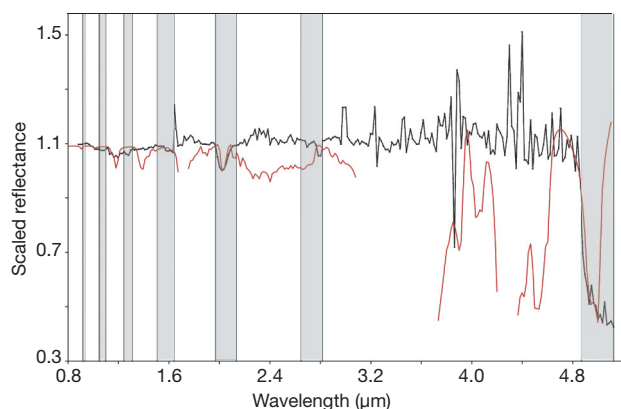
The  $I/F$  at 5  $\mu\text{m}$  of the central, dark interior of Ontario Lacus approaches zero when extrapolated to zero airmass, indicating that almost all photons incident on the lake interior at this wavelength are absorbed (Fig. 3). In contrast to the finite reflectivity of a nearby area outside the lake (Fig. 3), the lake can only have zero reflectivity if its surface is smooth and free of scattering centres greater than about 5  $\mu\text{m}$  in size, the surface is viewed in non-specular geometry, and there are no slopes or facets allowing partial specular reflection along the line of sight. It is highly unlikely that any solid surface in the lake interior could adhere to such constraints, thus indicating that Ontario Lacus is filled with a quiescent liquid, free of particles larger than a few micrometres.

In view of the widespread evidence for precipitation on Titan<sup>8–10</sup>, an obvious question is whether liquid methane exists in Ontario Lacus. Unfortunately, because methane gas is so abundant in Titan's atmosphere, and because the rotational–vibrational spectrum of liquid methane is nearly identical to that of methane gas (except for wavelength shifts that cannot be readily resolved in the VIMS data<sup>21</sup>), most features due to liquid methane would not be detectable in our spectra. Nevertheless, the steep decline in reflectance of Ontario Lacus within the 5- $\mu\text{m}$  window is consistent with the presence of other liquid alkanes such as ethane, propane and butane<sup>21–23</sup>.

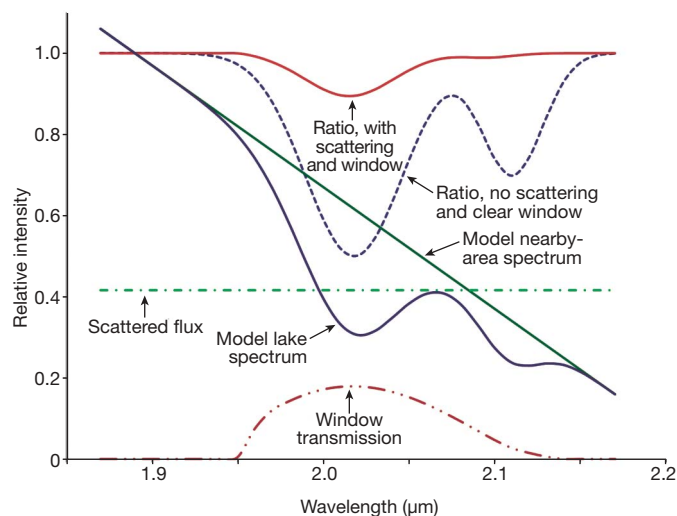
An expected alkane on Titan is liquid ethane<sup>1</sup>. In addition to many bands that lie outside Titan's atmospheric windows, model spectra of liquid ethane show three absorption bands within the 2- $\mu\text{m}$  atmospheric window (Fig. 4). Two of the bands are blended into one apparent band at VIMS resolution and are centred at 2.018  $\mu\text{m}$  (ref. 22). The third band is centred at 2.11  $\mu\text{m}$ , very near the long-wavelength limit of the 2- $\mu\text{m}$  window, and is weaker than the 2.018- $\mu\text{m}$  band. The 2.018- $\mu\text{m}$  band complex in our model spectrum matches the central wavelength and shape of that in the Ontario Lacus data, but is slightly narrower. The 2.11- $\mu\text{m}$  band is very weak in the Ontario Lacus data in comparison with that in the model spectrum, showing only as a small depression in the long-wavelength wing of the 2.018- $\mu\text{m}$  band.

The apparent differences between the model spectrum and the lake spectrum are explained when atmospheric scattering and opacity are considered. This is demonstrated by using a model spectrum of ethane calculated for the 2- $\mu\text{m}$  atmospheric window (Fig. 5) that includes the effects of atmospheric scattering and opacity. The simulated bands are clearly shallower, and for the strongest band they are broader because the band wings increase as the band depth increases. There is also a substantial decrease in the depth of the weaker band near the long-wavelength edge of the window, making the band complex seem to be a single band with the long-wavelength wing suppressed. The Ontario Lacus ratio spectrum (Fig. 2) clearly shows that the long-wavelength wing of the 2.018- $\mu\text{m}$  band is suppressed precisely as expected if it were an ethane band. We therefore take this as identification of an additional ethane band at 2.11  $\mu\text{m}$ .

Another strong indicator of lake composition is the steep decline in reflectance beyond 4.8  $\mu\text{m}$ . This is characteristic of alkanes in general, and ethane, propane and butane in particular<sup>21–23</sup>. The spectrum of liquid ethane matches the short-wavelength portion of the deep 5- $\mu\text{m}$  absorption complex in Titan's spectrum (Fig. 4), and possible narrow features near 4.94 and 5.00  $\mu\text{m}$ . Although the reflectance of liquid ethane turns steeply upwards near the middle of the 5- $\mu\text{m}$  window, a small amount of propane, butane and higher-order alkanes could easily explain the continued decline in Titan's reflectance there<sup>23</sup>. We therefore consider this to be additional strong evidence for the presence of liquid ethane in Ontario Lacus, with a suggestion of other compounds such as propane and butane. Finally, although our data cannot unequivocally determine the presence or absence of liquid methane in Ontario Lacus, recent work on the seasonal evolution of hydrocarbon lakes on Titan shows that if liquid ethane exists in these lakes, in all probability so does liquid methane<sup>24</sup>.



**Figure 4 | Ratio spectrum of Ontario's interior, with a model spectrum of liquid ethane superimposed.** The spectrum of liquid ethane is calculated for a path length of 300  $\mu\text{m}$  from tabulated absorption coefficients<sup>22</sup> and is convolved to the resolution and sampling interval of VIMS<sup>11</sup>. The black line shows the data, and the red line shows the model. The vertical grey bands denote the spectral channels within atmospheric windows through which VIMS can resolve surface detail.



**Figure 5 | Model spectrum of liquid ethane simulating the effects of aerosol scattering and atmospheric opacity.** The model uses two gaussians of the same depth and width as those seen in the ethane spectrum to simulate the effects of scattering. The plot shows model spectra for areas inside and outside the lake, along with their ratio spectrum. Also shown is a contribution from scattered light, a model window-transmission function (constructed from two gaussians appropriate for the methane-gas absorption bands on either side of Titan's 2- $\mu\text{m}$  window), and the effect that scattering and the window opacity have on the resultant spectrum. In addition to photons scattered from the surface through an atmospheric window, a substantial number of photons are scattered by aerosols whose scattering function is roughly constant across the window (or at least having no spectral features that mimic those originating from the surface). Consequently, decreases in reflected intensity resulting from surface-spectral-absorption features cannot reduce the observed total intensity below the atmospheric scattering floor. The result is that observed spectral features introduced by photons scattered from surface materials are made shallower and broader by the contribution of the scattered photons from aerosols, but the centre wavelengths of the bands are left unchanged (again, so long as the scattering intensity does not vary much across the narrow window, which is true of Titan's 2- $\mu\text{m}$  window). As can be clearly seen, the simulated bands are shallower and, especially for the strongest band, broader because the band wings increase as the band depth increases. Also demonstrated is the substantial reduction in depth of the weaker band near the long-wavelength edge of the window, making the band complex look like a single band with the long-wavelength wing suppressed.

There are other possible materials in Ontario Lacus that we have considered but have rejected on the basis of the existing data from Titan. One candidate is water ice. It has a strong 2- $\mu\text{m}$  absorption and is very absorbing in the 5- $\mu\text{m}$  region, but its 2- $\mu\text{m}$  band is much too broad and centred at a slightly shorter wavelength<sup>25,26</sup>. Ammonia and ammonia hydrate both have bands near 2  $\mu\text{m}$ , but they are too narrow and also centred at a slightly shorter wavelengths<sup>16,27</sup>. Carbon dioxide has a triplet centred near 2  $\mu\text{m}$ , but its components are too narrow and the complex is centred at a shorter wavelength<sup>16</sup>. Furthermore, all of the materials mentioned above are solid at Titan's temperatures and therefore none of them is likely to match the 5- $\mu\text{m}$  albedo of Ontario Lacus. There may be other simple compounds having bands in the 2- $\mu\text{m}$  region, but those discussed above are most plausible from a cosmochemical perspective<sup>28,29</sup>.

Received 28 February; accepted 19 May 2008.

1. Lunine, J. I., Stevenson, D. J. & Yung, Y. L. Ethane ocean on Titan. *Science* **222**, 1229–1230 (1983).
2. Lunine, J. I. Does Titan have an ocean—a review of current understanding of Titan's surface. *Rev. Geophys.* **31**, 133–149 (1993).
3. Lunine, J. I. Does Titan have oceans? *Am. Sci.* **82**, 134–143 (1994).
4. Dermott, S. F. & Sagan, C. Tidal effects of disconnected hydrocarbon seas on Titan. *Nature* **374**, 238–240 (1995).
5. Sagan, C. & Dermott, S. F. The tide in the seas of Titan. *Nature* **300**, 731–733 (1982).



6. Porco, C. C. *et al.* Imaging of Titan from the Cassini spacecraft. *Nature* **434**, 159–168 (2005).
7. Stofan, E. R. *et al.* The lakes of Titan. *Nature* **445**, 61–64 (2007).
8. Elachi, C. *et al.* Cassini radar views the surface of Titan. *Science* **308**, 970–974 (2005).
9. Elachi, C. *et al.* Titan radar mapper observations from Cassini's T-3 fly-by. *Nature* **441**, 709–713 (2006).
10. Tomasko, M. G. *et al.* Rain, winds and haze during the Huygens probe's descent to Titan's surface. *Nature* **438**, 765–778 (2005).
11. Brown, R. H. *et al.* The Cassini Visual and Infrared Mapping Spectrometer investigation. *Space Sci. Rev.* **115**, 111–168 (2004).
12. Griffith, C. A., Owen, T. & Wagener, R. Titan's surface and troposphere, investigated with ground-based, near-infrared observations. *Icarus* **93**, 362–378 (1991).
13. Griffith, C. A., Owen, T., Geballe, T. R., Rayner, J. & Rannou, P. Evidence for the exposure of water ice on Titan's surface. *Science* **300**, 628–630 (2003).
14. Smith, B. A. *et al.* A new look at the Saturn System—the Voyager-2 images. *Science* **215**, 504–537 (1982).
15. Smith, B. A. *et al.* Encounter with Saturn—Voyager-1 imaging science results. *Science* **212**, 163–191 (1981).
16. Brown, R. H. & Cruikshank, D. P. Determination of the composition and state of icy surfaces in the outer solar system. *Annu. Rev. Earth Planet. Sci.* **25**, 243–277 (1997).
17. Lopes, R. M. C. *et al.* The lakes and seas of Titan. *Eos* **88**, 569–570 (2007).
18. Barnes, J. W. *et al.* A 5-micron-bright spot on Titan: evidence for surface diversity. *Science* **310**, 92–95 (2005).
19. Soderblom, L. A. *et al.* Correlations between Cassini VIMS spectra and RADAR SAR images: Implications for Titan's surface composition and the character of the Huygens Probe landing site. *Planet. Space Sci.* **55**, 2025–2036 (2008).
20. Barnes, J. W. *et al.* Global-scale surface spectral variations on Titan seen from Cassini/VIMS. *Icarus* **186**, 242–258 (2007).
21. Grundy, W. M., Schmitt, B. & Quirico, E. The temperature-dependent spectrum of methane ice I between 0.7 and 5  $\mu\text{m}$  and opportunities for near-infrared remote thermometry. *Icarus* **155**, 486–496 (2002).
22. Clark, R. N. *et al.* Detection of widespread aromatic and aliphatic hydrocarbons on Titan. *Icarus* (in the press).
23. Clark, R. N., Curchin, J. M., Hoefen, T. M. & Swayze, G. A. Reflectance spectroscopy of organic compounds. I: Alkanes. *J. Geophys. Res.* (submitted).
24. Mitri, G., Showman, A. P., Lunine, J. I. & Lorenz, R. D. Hydrocarbon lakes on Titan. *Icarus* **186**, 385–394 (2007).
25. Clark, R. N. & Lucey, P. G. Spectral properties of ice–particulate mixtures: Implications for remote sensing. I: Intimate mixtures. *J. Geophys. Res.* **89**, 6341–6348 (1984).
26. Clark, R. N. & McCord, T. B. The Galilean satellites: New near-infrared reflectance measurements (0.65–2.5  $\mu\text{m}$ ) and a 0.325–4.08  $\mu\text{m}$  summary. *Icarus* **43**, 323–339 (1980).
27. Brown, R. H., Cruikshank, D. P., Tokunaga, A. T. & Smith, R. G. Search for volatiles on icy satellites. I: Europa. *Icarus* **74**, 262–271 (1988).
28. Prinn, R. G. & Fegley, B. Kinetic inhibition of CO and N<sub>2</sub> reduction in circumplanetary nebulae—implications for satellite composition. *Astrophys. J.* **249**, 308–317 (1981).
29. Lewis, J. S. Chemistry of planets. *Annu. Rev. Phys. Chem.* **24**, 339–351 (1973).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.H.B. (rhb@lpl.arizona.edu).

# Scaling the Kondo lattice

Yi-feng Yang<sup>1,2</sup>, Zachary Fisk<sup>3</sup>, Han-Oh Lee<sup>1</sup>, J. D. Thompson<sup>1</sup> & David Pines<sup>2</sup>

The origin of magnetic order in metals has two extremes: an instability in a liquid of local magnetic moments interacting through conduction electrons, and a spin-density wave instability in a Fermi liquid of itinerant electrons. This dichotomy between 'local-moment' magnetism and 'itinerant-electron' magnetism is reminiscent of the valence bond/molecular orbital dichotomy present in studies of chemical bonding. The class of heavy-electron intermetallic compounds of cerium, ytterbium and various 5f elements bridges the extremes, with itinerant-electron magnetic characteristics at low temperatures that grow out of a high-temperature local-moment state<sup>1</sup>. Describing this transition quantitatively has proved difficult, and one of the main unsolved problems is finding what determines the temperature scale for the evolution of this behaviour. Here we present a simple, semi-quantitative solution to this problem that provides a basic framework for interpreting the physics of heavy-electron materials and offers the prospect of a quantitative determination of the physical origin of their magnetic ordering and superconductivity. It also reveals the difference between the temperature scales that distinguish the conduction electrons' response to a single magnetic impurity and their response to a lattice of local moments, and provides an updated version of the well-known Doniach diagram<sup>2</sup>.

The physics of a single local-moment spin coupled to the conduction electrons in a simple metal—the single-ion Kondo problem—is now well understood and completely characterized by the strength,  $J$ , of that coupling and  $\rho$ , the density of electronic states of the host metal at the Fermi surface<sup>3</sup>. Until now no comparable understanding has been achieved for Kondo lattice materials—dense intermetallic compounds containing a chemically ordered lattice of local moments that are Kondo coupled to the compound's conduction electrons<sup>1</sup>. Such materials have low-temperature properties characterized by large effective electronic masses (whence the nomenclature 'heavy electron') deriving from collective local-moment deconfinement—a 4f or 5f local-moment–itinerant-electron transition that emerges below a characteristic temperature scale  $T^*$  (ref. 4). Above  $T^*$ , the local moments exhibit a Curie–Weiss-type magnetic susceptibility that begins to change at  $T^*$ . The deconfinement evolving below  $T^*$  becomes visible in Knight shift experiments<sup>5</sup> that show the emergence of a second component of the spin susceptibility, the heavy-electron Kondo liquid, whose logarithmic increase in density of states with decreasing temperature can be measured in these and other experiments<sup>6</sup>. It is generally accompanied by a drop in electrical resistivity, often called the development of coherence, as well as the development of a low-frequency Drude peak in the optical conductivity and signatures in other properties that we discuss in the Supplementary Information. The interplay of their single-site Kondo physics with intersite coupling effects gives the study of these intermetallic compounds a remarkable richness.

In the collective deconfinement, the entropy,  $R\ln 2$  (where  $R$  denotes the universal gas constant), carried by the local-moment degeneracy becomes associated with the itinerant-electron system.

This suggests that  $T^*$  is the temperature at which this entropy is developed in the lattice. Thus, with the entropy associated with the lattice vibrational degrees of freedom and excited crystal field contributions subtracted, we might find that  $S(T^*) \approx R\ln 2$ , assuming a ground-state crystal field doublet for the  $f$ -electrons, typical of cerium and ytterbium heavy-electron materials (other situations have a simple generalization). We find that this is observed experimentally: a magnetic entropy of approximately  $R\ln 2$  develops at temperature  $T^*$ , supporting the viewpoint that coherence represents the entanglement of the local-moment spin degree of freedom with those of the conduction electrons.

The unusual and interesting properties of Kondo lattice materials can thus be viewed as deriving from a large transfer of entropy from the high-temperature local-moment system, making  $T^*$  a primary parameter for characterizing such systems. We note that the heavy-electron superconductors in this class of materials typically develop an entropy of order  $(0.1\text{--}0.2)R\ln 2$  at the superconducting transition temperature,  $T_c$ , with  $T_c/T^*$  values of order  $1/20$ , suggesting a possible connection between  $T_c$  and  $T^*$ . It is of fundamental and practical importance to learn what determines the magnitude of  $T^*$ . We suggest a very simple solution to this problem: the Kondo coupling  $J$  governs both the single-ion screening (hybridization) that proceeds independently above  $T^*$  and the collective hybridization that produces coherent behaviour below  $T^*$ .

In a study of heavy-electron  $\text{Ce}_{1-x}\text{La}_x\text{CoIn}_5$  alloys, it has been argued that  $T^*$  corresponds to the nearest-neighbour intersite coupling<sup>7</sup>. Because this coupling may reasonably be expected to arise primarily from the coupling of adjacent cerium ions through the conduction electrons (namely the Ruderman–Kittel–Kasuya–Yosida (RKKY) interaction), which increases as  $k_B T_{\text{RKKY}} \sim J^2 \rho$  (where  $k_B$  denotes Boltzmann's constant and  $T_{\text{RKKY}}$  the RKKY temperature), on general grounds we would expect  $T^*$  to take the form<sup>1</sup>

$$T^* = cJ^2 \rho \quad (1)$$

where  $J$  is the local Kondo coupling,  $\rho$  is the density of states of the conduction electrons coupled to the local spins and  $c$  is a constant determined by the details of the hybridization and the conduction electron Fermi surface. Here and in the following we set  $k_B = 1$  for simplicity. The density of states  $\rho$  can be estimated from the Sommerfeld coefficient  $\gamma$  of the corresponding non-magnetic lanthanum (or yttrium, lutetium or thorium) compounds<sup>3</sup>:

$$\rho = \frac{3\gamma}{\pi^2}$$

On the other hand, the single-ion Kondo temperature,  $T_K$ , is given approximately by<sup>3</sup>:

$$T_K = \rho^{-1} e^{-1/J\rho}$$

Because both temperature scales are determined by the single-ion Kondo coupling  $J$ , we have the following equation:

<sup>1</sup>Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. <sup>2</sup>Department of Physics, University of California, Davis, California 95616, USA. <sup>3</sup>Department of Physics and Astronomy, University of California, Irvine, California 92697, USA.

$$J\rho = -1/\ln(T_K\rho) = \sqrt{c^{-1}T^*\rho} \quad (2)$$

From a plot of  $\sqrt{T^*\rho}$  against  $-1/\ln(T_K\rho)$  it is therefore possible to determine whether equation (1) provides a useful expression for  $T^*$ , that is, one in which the value of  $c$  changes little from one material to another.

We have accordingly examined a wide variety of heavy-electron materials for which both  $T^*$  and  $T_K$  are already known or can be determined, with the results presented in Table 1. In constructing Table 1, we used the fact that the deconfinement of the local moments has a number of direct physical consequences, any of which can be used to estimate  $T^*$ : the evolution of the magnetic entropy towards  $R\ln 2$ , the development of a low-frequency Drude peak in the optical conductivity, the onset of a Knight shift or Hall anomaly, the departure of the magnetic susceptibility from Curie–Weiss behaviour, and a coherence peak in the magnetic resistivity. These different methods gave roughly the same results as long as other physical effects are not present or can be taken into account. We note that, in general, crystal field effects may have a role, so the actual behaviour may be more complicated than a simple picture based on the dominance of  $T^*$  would indicate. Determining  $T_K$  also requires some care. The most straightforward way to determine  $T_K$  for a given material is to use measurements of single-ion Kondo behaviour in the corresponding dilute compounds with a very low cerium (or uranium or ytterbium) concentration in the non-magnetic host. Because  $J$  and  $\rho$  are volume dependent, differences between the unit-cell volumes of the non-magnetic reference host and the corresponding fully dense Kondo lattice must be taken into account. As discussed in the Supplementary Information, an equivalent volume can be achieved either by applying external pressure or by proper chemical substitution.

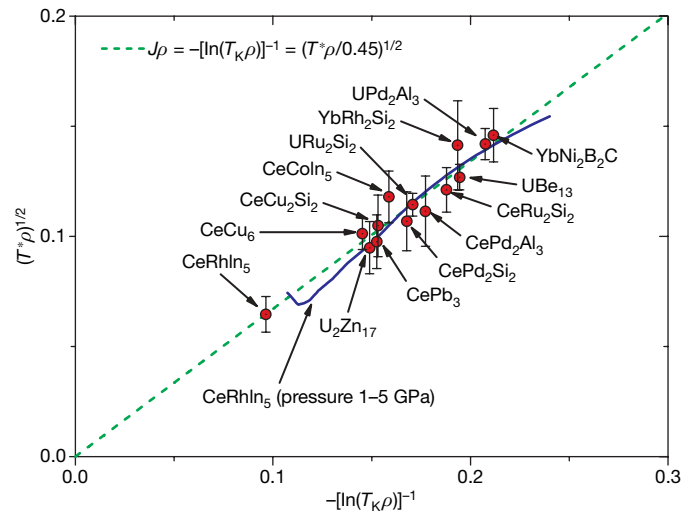
Using the results for  $T_K$  and  $T^*$  given in Table 1, we obtain the result shown in Fig. 1, where a choice of  $c \approx 0.45$  is seen to provide an excellent account of  $T^*$  for a broad range of materials that have cubic, tetragonal or hexagonal crystal structures and a magnetically ordered, superconducting or paramagnetic ground state. It is remarkable that a simple relation between  $T^*$ , the single-ion Kondo scale  $T_K$ , and the unenhanced density of states  $\rho$  should be semi-quantitatively valid for such a variety of materials. Significant variation in the material-specific factor,  $c$ , is expected, yet this seems a weak effect for this set of materials. Moreover, the dominance of single-ion, local-moment compensation when  $T_K > T_{\text{RKKY}}$ , which was expected to occur frequently in heavy-electron materials, is not present in this group of materials, although it could still appear in a larger data set. The present compilation suggests that there is a single dominant energy scale in the physics of heavy-electron materials—namely the Kondo coupling  $J$ —that determines both  $T_K$  and  $T^*$ . We note that the value of the structure factor  $c$  will be slightly different if a different formula is used for  $T_K$ , but the overall behaviour is not affected.

**Table 1 | Experimental  $T^*$ ,  $T_K$  and  $\gamma$  values for a variety of Kondo lattice compounds**

| Compound                           | $T^*$ (K) | $T_K$ (K) | $\gamma$ (mJ mol <sup>-1</sup> K <sup>-2</sup> ) | $J\rho$ | $J$ (meV) | $c$  | Reference      |
|------------------------------------|-----------|-----------|--|---------|-----------|------|----------------|
| CeRhIn <sub>5</sub>                | 20 ± 5    | 0.15      | 5.7  | 0.10    | 40        | 0.45 | 6, 8, H.-O.L.* |
| CeCu <sub>6</sub>                  | 35 ± 5    | 3.5       | 8  | 0.15    | 43        | 0.49 | 9, 10          |
| CeCu <sub>2</sub> Si <sub>2</sub>  | 75 ± 20   | 10        | 4  | 0.15    | 90        | 0.47 | 6, 11, 12      |
| CePb <sub>3</sub>                  | 20 ± 5    | 3         | 13   | 0.15    | 28        | 0.41 | 13, 14         |
| CeCoIn <sub>5</sub>                | 50 ± 10   | 6.6       | 7.6  | 0.16    | 49        | 0.55 | 4, 6, 7        |
| CePd <sub>2</sub> Si <sub>2</sub>  | 40 ± 10   | 9         | 7.8  | 0.17    | 51        | 0.41 | 15, 16         |
| CePd <sub>2</sub> Al <sub>3</sub>  | 35 ± 10   | 10        | 9.7  | 0.18    | 43        | 0.40 | 17, 18, 19     |
| CeRu <sub>2</sub> Si <sub>2</sub>  | 60 ± 10   | 20        | 6.68   | 0.19    | 66        | 0.42 | 20, 21         |
| U <sub>2</sub> Zn <sub>17</sub>    | 20 ± 5    | 2.7       | 12.3   | 0.15    | 29        | 0.41 | 22, 23         |
| URu <sub>2</sub> Si <sub>2</sub>   | 55 ± 5    | 12        | 6.5  | 0.17    | 62        | 0.45 | 6, 24, 25      |
| UBe <sub>13</sub>                  | 55 ± 5    | 20        | 8  | 0.19    | 57        | 0.43 | 26, 27         |
| UPd <sub>2</sub> Al <sub>3</sub>   | 60 ± 10   | 25        | 9.7  | 0.21    | 51        | 0.48 | 19, 28         |
| YbRh <sub>2</sub> Si <sub>2</sub>  | 70 ± 20   | 20        | 7.8  | 0.19    | 58        | 0.53 | Z.F.†          |
| YbNi <sub>2</sub> B <sub>2</sub> C | 50 ± 5    | 20        | 11   | 0.21    | 44        | 0.47 | 29             |

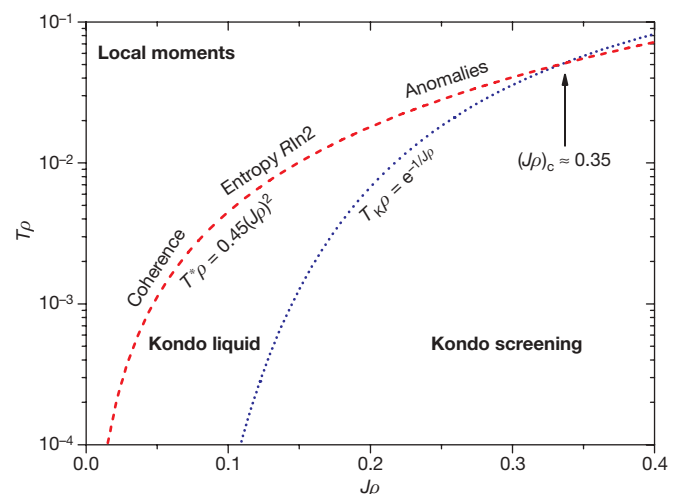
\*H.-O. Lee, unpublished resistivity and specific heat measurements on Ce<sub>1-x</sub>La<sub>x</sub>RhIn<sub>5</sub>.

†Z. Fisk, unpublished resistivity and specific heat measurements on Yb<sub>1-x</sub>Lu<sub>x</sub>Rh<sub>2</sub>Si<sub>2</sub>.



**Figure 1 | Confirmation of  $T^*$  given by the intersite RKKY interaction for a variety of Kondo lattice materials.** The solid line shows  $T^*$  (resistivity peak) of CeRhIn<sub>5</sub> under pressure from 1 GPa (lower left) to 5 GPa (upper right). Each error bar indicates a typical range of uncertainties for estimating  $T^*$  (see the Supplementary Information). The dashed line is a guide to all the materials. The evident correlation between  $T^*$ ,  $T_K$  and  $\rho$  provides verification of equation (2) and the RKKY nature of  $T^*$ . The small deviation from the  $c = 0.45$  line may result from details of the local hybridization and the conduction electron Fermi surface.

Our approach also enables us to update the well-known Doniach diagram<sup>2</sup>, using our experimental values of  $T^*$  in place of estimates of  $T_{\text{RKKY}}$  and comparing their variation with  $J$  to that of  $T_K$  for a specific choice of  $c$ . The result is shown in Fig. 2. Remarkably, for  $c = 0.45$ , we find that  $T^* = T_K$  for  $J\rho \approx 0.35$ , a value very close to that obtained for the one-dimensional Kondo necklace<sup>30</sup>. We call attention to the fact that the heavy-electron materials considered here that appear to exhibit quantum critical behaviour have  $J$  values such that  $J\rho$  clusters between 0.15 and 0.20, values that are low in comparison with those (0.35 or greater) at which the single-ion Kondo coupling may be expected to play a significant role. Furthermore, the cerium-based heavy-electron superconductors CeCu<sub>2</sub>Si<sub>2</sub> and CeCoIn<sub>5</sub>, as well as



**Figure 2 | Updated Doniach diagram for Kondo lattice materials.** The dashed line indicates the intersite RKKY temperature,  $T^*$ , with  $c = 0.45$ . Below  $T^*$ , the local 4f (or 5f) moments become partially itinerant, giving rise to the universal Kondo liquid behaviour and rich phase diagrams. The dotted line is the single-ion Kondo temperature  $T_K$ , below which the remaining local moments are screened.  $T_K$  reaches  $T^*$  at the critical value  $(J\rho)_c \approx 0.35$ .



CeRhIn<sub>5</sub> and CePd<sub>2</sub>Si<sub>2</sub> under pressure, cluster even more tightly near  $J\rho = 0.16$ , implying the possibility of a necessary but not sufficient condition on where new examples might be found.

Our results suggest an answer to an interesting piece of the problem of the dense Kondo lattice, which, by inference, provides a route to similar organizing principles in correlated electron materials more broadly defined. The dense Kondo materials are in many ways the simplest highly correlated electron materials, as charge fluctuations are unimportant. In moving from 4f- to 5f-electron materials, we encounter new low-temperature ground states that condense from the heavy-electron fluid and whose nature remains a mystery. A first step into the physics of these highly correlated electron materials in which charge fluctuations come into play may be the study of how  $T^*$  evolves from the simple limiting behaviour described here. Progress in understanding these materials should apply analogously to strongly correlated 3d-electron systems, whose physics is in many respects that of 5f-electron systems but with correspondingly higher energy scales.

## METHODS SUMMARY

For a broad spectrum of heavy-electron materials, we determine three fundamental physical quantities: the local Kondo coupling,  $J$ , between a single local moment and the background conduction electrons, and the two distinct temperature scales,  $T_K$  and  $T^*$ , that characterize the magnetic response of the conduction electrons to the introduction of local moments.  $T_K$  is the Kondo temperature that governs their magnetic response to a single magnetic impurity, and  $T^*$  is the coherence temperature that governs their collective response to a lattice of local moments.

$T_K$  is usually estimated from measurements on dilute impurities in an appropriate non-magnetic host material that typically has lattice parameters which are different from those of the dense Kondo lattice. It is therefore necessary to apply pressure or make chemical substitutions to arrive at isovolumetric values for  $T_K$  and  $J$  in the Kondo lattice material. A knowledge of  $T_K$  and the electronic density of states,  $\rho$ , which is obtained from the Sommerfeld coefficient,  $\gamma$ , of the host material, then makes it possible to determine  $J$ .

The emergence of the Kondo liquid at  $T^*$  that is accompanied by a loss in strength of the local moments leads to observable physical effects in a number of phenomena. The determination of  $T^*$  from these effects, including changes in the optical conductivity, the magnetic entropy, the magnetic resistivity, and point-contact spectroscopy, as well as from the emergence of a Knight shift anomaly, is described in detail in the Supplementary Information.

Received 7 December 2007; accepted 5 June 2008.

1. von Löhneysen, H., Rosch, A., Vojta, M. & Wölfle, P. Fermi-liquid instabilities at magnetic quantum phase transitions. *Rev. Mod. Phys.* **79**, 1015–1075 (2007).
2. Doniach, S. The Kondo lattice and weak antiferromagnetism. *Physica B* **91**, 231–234 (1977).
3. Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, 1993).
4. Nakatsuji, S., Pines, D. & Fisk, Z. Two fluid description of the Kondo lattice. *Phys. Rev. Lett.* **92**, 016401 (2004).
5. Curro, N. J., Young, B.-L., Schmalian, J. & Pines, D. Scaling in the emergent behavior of heavy-electron materials. *Phys. Rev. B* **70**, 235117 (2004).
6. Yang, Y.-F. & Pines, D. Universal behavior in heavy-electron materials. *Phys. Rev. Lett.* **100**, 096404 (2008).
7. Nakatsuji, S. *et al.* Intersite coupling effects in a Kondo lattice. *Phys. Rev. Lett.* **89**, 106402 (2002).
8. Hegger, H. *et al.* Pressure-induced superconductivity in quasi-2D CeRhIn<sub>5</sub>. *Phys. Rev. Lett.* **84**, 4986–4989 (2000).
9. Marabelli, F. & Wachter, P. Temperature dependence of the optical conductivity of the heavy-fermion system CeCu<sub>6</sub>. *Phys. Rev. B* **42**, 3307–3311 (1990).
10. Satoh, K., Fujita, T., Maeno, Y., Ōnuki, Y. & Komatsubara, T. Low-temperature specific heat of Ce<sub>x</sub>La<sub>1-x</sub>Cu<sub>6</sub>. *J. Phys. Soc. Jpn* **58**, 1012–1020 (1989).

11. Aliev, F. G., Brandt, N. B., Moshchalkov, V. V. & Chudinov, S. M. Electric and magnetic properties of the Kondo-lattice compound CeCu<sub>2</sub>Si<sub>2</sub>. *J. Low Temp. Phys.* **57**, 61–93 (1984).
12. Goremychkin, E. A., Osborn, R. & Muzychka, A. Yu. Crystal-field effects in PrCu<sub>2</sub>Si<sub>2</sub>: An evaluation of evidence for heavy-fermion behavior. *Phys. Rev. B* **50**, 13863–13866 (1994).
13. Dürkop, D. *et al.* Antiferromagnetic order and 4f-instability of CePb<sub>3</sub>. *Z. Phys. B* **63**, 55–61 (1986).
14. Lin, C. L., Wallash, A., Crow, J. E., Mihalisin, T. & Schlottmann, P. Heavy-fermion behavior and the single-ion Kondo model. *Phys. Rev. Lett.* **58**, 1232–1235 (1987).
15. Besnus, M. J., Braghta, A. & Meyer, A. Kondo behaviour in magnetic (Ce-La)Pd<sub>2</sub>Si<sub>2</sub>. *Z. Phys. B* **83**, 207–211 (1991).
16. Palstra, T. T. M. *et al.* Superconductivity in the ternary rare-earth (Y, La, and Lu) compounds RPd<sub>2</sub>Si<sub>2</sub> and RRh<sub>2</sub>Si<sub>2</sub>. *Phys. Rev. B* **34**, 4566–4570 (1986).
17. Ghosh, K., Ramakrishnan, S., Malik, S. K. & Chandra, G. Resistivity and magnetic-susceptibility studies in the RPd<sub>2</sub>Al<sub>3</sub> (R=La, Ce, Pr, Nd, and Sm) system. *Phys. Rev. B* **48**, 6249–6254 (1993).
18. Medina, A. N., Rojas, D. P., Gandra, F. G., Azanha, W. R. & Cardoso, L. P. Change of the Kondo regime in CePd<sub>2</sub>Al<sub>3</sub> induced by chemical substitution: Verification of the Doniach diagram. *Phys. Rev. B* **59**, 8738–8744 (1999).
19. Zapf, V. S., Dickey, R. P., Freeman, E. J., Sirvent, C. & Maple, M. B. Magnetic and non-Fermi-liquid properties of U<sub>1-x</sub>La<sub>x</sub>Pd<sub>2</sub>Al<sub>3</sub>. *Phys. Rev. B* **65**, 024437 (2001).
20. Kitaoka, Y., Arimoto, H., Kohori, Y. & Asayama, K. NMR study of the dense Kondo compound CeRu<sub>2</sub>Si<sub>2</sub>. *J. Phys. Soc. Jpn* **54**, 3236–3239 (1985).
21. Matsuhira, K. *et al.* Single-site and inter-site effects in heavy fermion compound CeRu<sub>2</sub>Si<sub>2</sub> studied by constant volume dilution. *J. Phys. Soc. Jpn* **66**, 2851–2863 (1997).
22. Ott, H. R., Rudigier, H., Delsing, P. & Fisk, Z. Magnetic ground state of a heavy-electron system: U<sub>2</sub>Zn<sub>17</sub>. *Phys. Rev. Lett.* **52**, 1551–1554 (1984).
23. Takagi, S., Suzuki, H. & Anzai, K. Kondo effect of U impurities in dilute (YU)<sub>2</sub>Zn<sub>17</sub>. *J. Phys. Soc. Jpn* **70**, 3098–3106 (2001).
24. Yokoyama, M., Amitsuka, H., Kuwahara, K., Tenya, K. & Sakakibara, T. Anomalous Fermi liquid behavior of the dilute uranium alloys La<sub>1-x</sub>U<sub>x</sub>Ru<sub>2</sub>Si<sub>2</sub> ( $x \leq 0.07$ ). *J. Phys. Soc. Jpn* **71**, 3037–3042 (2002).
25. van der Meulen, H. P. *et al.* Field suppression of the heavy-fermion state in CeRu<sub>2</sub>Si<sub>2</sub>. *Phys. Rev. B* **44**, 814–818 (1991).
26. Sonier, J. E. *et al.*  $\mu^+$  Knight shift measurements in U<sub>0.965</sub>Th<sub>0.035</sub>Be<sub>13</sub> single crystals. *Phys. Rev. Lett.* **85**, 2821–2824 (2000).
27. Kim, J. S., Andraka, B., Jee, C. S., Roy, S. B. & Stewart, G. R. Single-ion effects in the formation of the heavy-fermion ground state in UBe<sub>13</sub>. *Phys. Rev. B* **41**, 11073–11081 (1990).
28. Kyogaku, M. *et al.* NMR and NQR studies of magnetism and superconductivity in the antiferromagnetic heavy fermion superconductors UM<sub>2</sub>Al<sub>3</sub> (M=Ni and Pd). *J. Phys. Soc. Jpn* **62**, 4016–4030 (1993).
29. Li, S. *et al.* Physical properties of Lu<sub>1-x</sub>Yb<sub>x</sub>Ni<sub>2</sub>B<sub>2</sub>C. *Phil. Mag.* **86**, 3021–3041 (2006).
30. Jullien, R., Fields, J. N. & Doniach, S. Zero-temperature real-space renormalization-group method for a Kondo-lattice model Hamiltonian. *Phys. Rev. B* **16**, 4889–4900 (1977).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank our colleagues at the August 2007 ICAM workshop on 1-1-5 materials, where this work originated, for many useful discussions and V. Sidorov for sharing unpublished measurements of the resistivities of LaRhIn<sub>5</sub> and La<sub>0.975</sub>Ce<sub>0.025</sub>RhIn<sub>5</sub> at pressures to 5 GPa. Y.Y. wishes to thank ICAM for the fellowship that has made this collaboration possible. Z.F. was supported by NSF grant NSF-DMR-0710492. D.P. acknowledges support from start-up funding from the Physics Department of the University of California, Davis. Work at Los Alamos was performed under the auspices of the US Department of Energy, Office of Science, and supported in part by the Los Alamos Directed Research and Development program.

**Author Contributions** The data analysis is primarily due to Y.Y., with some contributions from D.P., who with Z.F. developed some of the basic physical ideas tested here. Experiments on Ce<sub>1-x</sub>La<sub>x</sub>RhIn<sub>5</sub> under pressure were carried out by H.-O.L. and J.D.T.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Y.Y. (yifengyyf@gmail.com).

## LETTERS

# Calendars with Olympiad display and eclipse prediction on the Antikythera Mechanism

Tony Freeth<sup>1,2</sup>, Alexander Jones<sup>3</sup>, John M. Steele<sup>4</sup> & Yanis Bitsakis<sup>1,5</sup>

Previous research on the Antikythera Mechanism established a highly complex ancient Greek geared mechanism with front and back output dials<sup>1–7</sup>. The upper back dial is a 19-year calendar, based on the Metonic cycle, arranged as a five-turn spiral<sup>1,6,8</sup>. The lower back dial is a Saros eclipse-prediction dial, arranged as a four-turn spiral of 223 lunar months, with glyphs indicating eclipse predictions<sup>6</sup>. Here we add surprising findings concerning these back dials. Though no month names on the Metonic calendar were previously known, we have now identified all 12 months, which are unexpectedly of Corinthian origin. The Corinthian colonies of northwestern Greece or Syracuse in Sicily are leading contenders—the latter suggesting a heritage going back to Archimedes. Calendars with excluded days to regulate month lengths, described in a first century BC source<sup>9</sup>, have hitherto been dismissed as implausible<sup>10,11</sup>. We demonstrate their existence in the Antikythera calendar, and in the process establish why the Metonic dial has five turns. The upper subsidiary dial is not a 76-year Callippic dial as previously thought<sup>8</sup>, but follows the four-year cycle of the Olympiad and its associated Panhellenic Games. Newly identified index letters in each glyph on the Saros dial show that a previous reconstruction needs modification<sup>6</sup>. We explore models for generating the unusual glyph distribution, and show how the eclipse times appear to be contradictory. We explain the four turns of the Saros dial in terms of the full moon cycle and the Exeligmos dial as indicating a necessary correction to the predicted eclipse times. The new results on the Metonic calendar, Olympiad dial and eclipse prediction link the cycles of human institutions with the celestial cycles embedded in the Mechanism's gearwork.

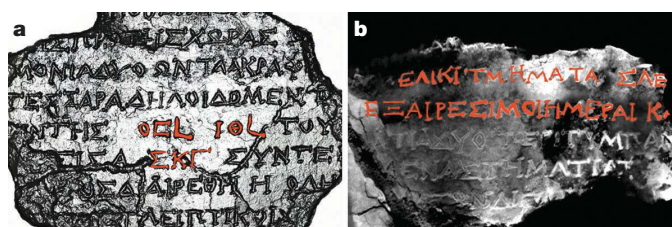
This extraordinary astronomical mechanism from about 100 BC employed bronze gears to make calculations based on cycles of the Solar System<sup>1,6</sup> (Supplementary Notes 1). Recovered in 1901 by Greek sponge-divers, its corroded remains are now split into 82 fragments—7 larger fragments (A–G) and 75 smaller fragments (1–75)<sup>6</sup>. Data, gathered in 2005<sup>6,7</sup>, included still photography, digital surface imaging<sup>12</sup> and, crucially for this study, microfocus X-ray computed tomography (CT)<sup>6,13</sup> (Figs 1–3)—details are in Supplementary Notes 2 (and at [www.antikythera-mechanism.gr](http://www.antikythera-mechanism.gr)).

The main upper back dial is now established as a Metonic calendar<sup>1,6,8</sup> (Figs 1 and 2, Supplementary Box 1). The calendar dial bears inscriptions, only viewable using X-ray CT. We have now identified all 12 months of this calendar (Fig. 2, Supplementary Notes 3), providing conclusive evidence of the regulation of a Greek civil calendar by a Metonic cycle, and clues to the instrument's origin. Whereas the Babylonian calendar followed a Metonic cycle from about 500 BC, it has commonly been assumed that the intercalary months of the numerous lunisolar calendars of the Greek cities were determined arbitrarily—Metonic and Callippic cycles

(Supplementary Box 1) only being used by astronomers<sup>14</sup>. The month names on the Metonic spiral, however, belong to a regional calendar unassociated with technical astronomy, suggesting that it may have been common for Greek civil calendars to follow the Metonic cycle by about 100 BC.

The inscriptions show that not only the names and order of the months were regulated, but also which years had 13 months, which month was repeated in these years, and which months had 29 or 30 days. The rules are similar to those given by the first century BC writer Geminus<sup>9</sup>, whose accuracy has hitherto been in doubt<sup>10,11</sup>. Years are numbered 1 to 19, and intercalary months are spread as evenly as possible over the cycle, such that each year begins with the first new moon following solstice or equinox<sup>15</sup>. In a Metonic cycle, 110 of the 235 months must have 29 days (Supplementary Box 1). The divisibility of both 110 and 235 by 5 explains the five turns of the spiral: months on the same radius across all five turns are equal in length. The numbers on the inside of each 29-day radius indicate which day in these months is skipped (Fig. 2). The skipped days are spread uniformly at intervals of 64 or 65 days across successive Metonic periods, improving on Geminus' scheme, which had uniform 64-day intervals followed by a run of 74 unskipped days at the end.

The month names and order in Greek regional calendars vary widely<sup>16</sup>. The months on the Mechanism belong to one of the Dorian family of calendars, with practically a complete match (11 or 12 names) with Illyria and Epirus in northwestern Greece and with Corcyra (Corfu)—all Corinthian colonies. The calendars of Corinth and its other important colonial foundation, Syracuse, are poorly



**Figure 1 | The 'instruction manual'.** Previously identified inscriptions<sup>1,6</sup> reveal remnants of an instruction manual, describing the Mechanism's cycles, dials and functions, as seen in two examples from the Mechanism's back door. **a**, Polynomial texture mapping of fragment 19 shows fine surface detail, with text about 2 mm high. Highlighted in red are "76 years, 19 years" for the Callippic and Metonic cycles (Supplementary Box 1), and "223", for the Saros cycle (Supplementary Box 2). **b**, X-ray CT of fragment E reveals text about 2 mm high. Highlighted are "on the spiral subdivisions 235", confirming the Metonic dial (Supplementary Box 1); and "excluded days 2...", the final "K" presumably standing for the number 20—part of the 22 excluded days round each of the five turns of the Metonic calendar—though "B" that would complete "KB" (22) remains speculative.

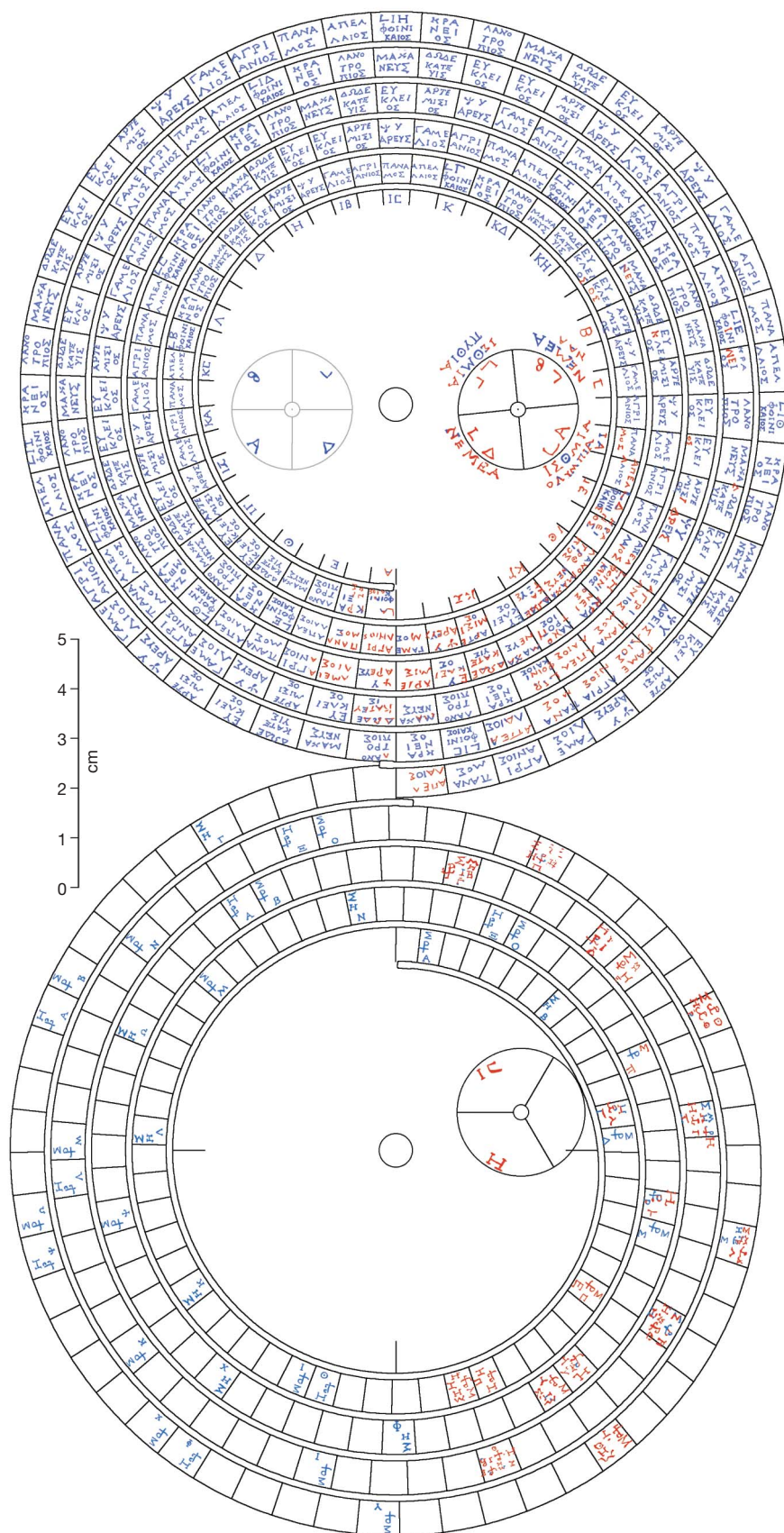
<sup>1</sup>Antikythera Mechanism Research Project, 3 Tyrwhitt Crescent, Roath Park, Cardiff CF23 5QP, UK. <sup>2</sup>Images First Ltd, 10 Hereford Road, South Ealing, London W5 4SE, UK. <sup>3</sup>Institute for the Study of the Ancient World, 15 East 84th Street, New York, New York 10028, USA. <sup>4</sup>Department of Physics, University of Durham, Rochester Building, South Road, Durham DH1 3LE, UK. <sup>5</sup>Centre for History and Palaeography, 3, P. Skouze str., GR-10560 Athens, Greece.



documented. Seven of the Mechanism's months, however, coincide in both name and sequence with the calendar of Tauromenion in Sicily, which was probably originated by settlers from Syracuse in the

fourth century BC. The Mechanism's calendar is thus from Corinth or one of its colonies. Moreover, the estimated date of the Mechanism falls after the Roman devastation of Corinth (146 BC) and Epirus

**Figure 2 | The back dials.** Text in red is traced from X-ray CT; text in blue is reconstructed. Top, the Metonic dial is the main upper dial: a 19-year calendar with 235 months round a five-turn spiral. Though the evidence is scant, we have fortunately been able to decipher all the month names because of their repetition round the dial. With reasonable assumptions about which years have 13 months and which months are repeated in these years, we can then reconstruct the whole of the calendar because of its cyclical nature. The newly identified Corinthian months, written over two or three lines in each cell, are: 1, ΦΟΙΝΙΚΑΙΟΣ; 2, ΚΡΑΝΕΙΟΣ; 3, ΛΑΝΟΤΡΟΠΙΟΣ; 4, ΜΑΧΑΝΕΥΣ; 5, ΔΩΔΕΚΑΤΕΥΣ; 6, ΕΥΚΑΕΙΟΣ; 7, ΑΡΤΕΜΙΣΙΟΣ; 8, ΨΥΔΡΕΥΣ; 9, ΓΑΜΕΙΛΙΟΣ; 10, ΑΓΓΙΑΝΙΟΣ; 11, ΠΑΝΑΜΟΣ; 12, ΑΠΕΛΛΑΙΟΣ. The numbers A (1), E (5), Θ (9), ΙΓ (13)... around the inside of the spiral specify the excluded days to be skipped in each of the five 29-day months on the same radius. Within the Metonic dial are shown two subsidiary dials. Right, the Olympiad dial (see Fig. 3), which is identified here for the first time. It is a four-year dial, representing the cycle of the Panhellenic Games, a central part of ancient Greek culture and a common basis for chronology. Left, the hypothetical Callippic dial, which follows a 76-year cycle, indicated on the back door inscriptions (Fig. 1). Bottom, the Saros dial is the main lower dial: an 18-year (223-lunar month) scale over a four-turn spiral, for predicting eclipses. Predictions are shown in the relevant months as glyphs (see Fig. 4), which indicate lunar and solar eclipses and their predicted times of day. This new reconstruction has 51 glyphs, specifying 38 lunar and 27 solar eclipses. The glyph times are incomplete as their generation remains obscure. The divisions on the inside of the dial at the cardinal points indicate the start of a new full moon cycle (Supplementary Box 2). Within the Saros dial is shown a subsidiary dial, the Exeligmos dial: this is a 54-year triple Saros dial, whose function is now understood. The first sector is blank (representing 0) and the following are labelled with numbers H (8) and Ιζ (16). The dial pointer indicates which number must be added to the glyph times in hours to get the eclipse times.





(171–168 BC). Syracuse's candidacy suggests a possible mechanical tradition going back to Archimedes (died 212 BC), who invented a planetarium described by Cicero<sup>17</sup> (first century BC) and wrote a lost book on astronomical mechanisms<sup>18</sup>.

The subsidiary dial (Fig. 3) inside the Metonic spiral was formerly believed to be a 76-year Callippic dial<sup>18</sup> (Supplementary Box 1). We have now established from its inscriptions that it displays the 4-year Olympiad cycle—a suggestion made previously for the main upper back dial<sup>19</sup>. The four sectors are inscribed anticlockwise with each sector containing a year number and two Panhellenic Games: the 'crown' games of Isthmia, Olympia, Nemea and Pythia and two lesser games: Naa (at Dodona) and a second game not yet deciphered<sup>20,21</sup>. As biennial games, Isthmia and Nemea occur twice. The Olympiads were a common framework for chronology, with years normally beginning in midsummer. But here the year must start between early autumn and early spring, because the Isthmian Games are in the years preceding their usual positions in the cycle (Fig. 3). Several month names favour a start following the autumnal equinox. The small ( $\sim 8^\circ$ , that is, one month) offset of the dial took account of the variation in the start of the lunisolar calendar and so ensured that the next Olympiad year would never start before the current year's games were over.

The Olympiad dial must be turned from the existing gearing<sup>6</sup> at a rate of one-quarter turn per year. Underneath the Olympiad dial are the remains of an isolated gear with 60 teeth<sup>1,6</sup>. Engaging this with a single additional gear with 57 teeth on the shaft of the Metonic pointer provides the correct anticlockwise rotation. Sizing this gear, with tooth pitch equal to the 60-tooth gear, gives a gear radius exactly as required by the interaxial distance: strong supporting evidence both for the Olympiad dial and this mechanical arrangement. The "76 years" inscription (Fig. 1) and other factors favour a Callippic dial, as a second subsidiary, symmetrical with the Olympiad dial—though loss of evidence means confirmation is unlikely. Might a fourth subsidiary, symmetric with the Exeligmos dial (Figs 1 and 2) complete the dial system? An existing shaft here does not penetrate

the back plate and does not appear to rotate at any meaningful rate. So this seems doubtful.

We have increased the number of identified eclipse glyphs<sup>6</sup> (Figure 4, Supplementary Notes 4) from 16 to 18. The decoding of these glyphs is extended here with the observation of 'N<sup>Y</sup>' abbreviating 'N<sup>Y</sup>KTOΣ' ('of the night') for solar eclipses and index letters at the bottom of each glyph in alphabetical order. These mean that the Saros dial starts at the top (as initially suggested<sup>4</sup>) with index letter 'A' rather than at the bottom (as in a later model<sup>6</sup>). With any other dial start, extrapolation of the index letters back to the first glyph would force them to begin in the middle of the alphabet. The alphabetic index letters also constrain the number of glyphs. If the 18 glyphs are aligned with lunar and solar eclipses in the last four centuries BC<sup>22</sup>, they give a perfect match for 100 start dates, suggesting an excellent prediction scheme. However, we do not believe that the glyphs were based primarily on observations: even extensive observations over decades would miss a high proportion of the 65 estimated eclipse predictions (Supplementary Box 2). The glyphs appear to have been generated by a scheme of eclipse possibilities, similar to Babylonian schemes with an 8-7-8-7-8-pattern<sup>23,24</sup> (Supplementary Box 2), which can be generated from a simple arithmetical model of nodal elongation at syzygy<sup>25</sup>. However, these schemes have 5- or 6-month gaps between all predictions, whereas the index letters imply that the Antikythera scheme has some longer gaps.

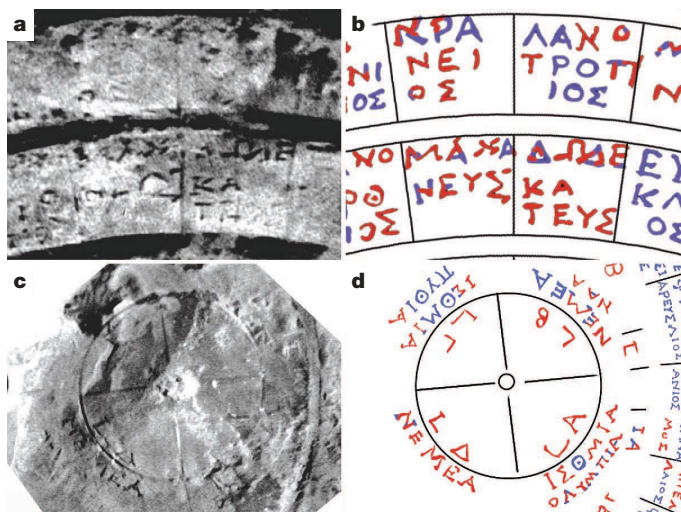
We consider kinematic models for glyph generation, defined by different nodal elongation criteria and computable either with similar technology to the Mechanism or an arithmetic method. With suitable criteria, these models generate all the Antikythera glyphs and no glyphs which are observably absent—whether the model uses mean months or first anomaly months (incorporating lunar and solar anomalies). However, none of these models exactly match the index letters. Because of parallax, the likelihood of a solar eclipse depends not only on a syzygy's nodal elongation, but also on whether it occurs north or south of the ecliptic, as was recognized in antiquity<sup>11</sup>. Introducing this asymmetry produces models (both using mean and first anomaly months) that exactly match all 18 definite glyphs with a single index letter error, caused by the only instance where the models generate two adjacent lunar glyphs—something which never featured in Babylonian schemes. Discarding the second of these gives a perfect match. These kinematic models provide a persuasive explanation of the glyph sequence.

Matching the glyph times with actual eclipse times over the last four centuries BC has not discovered close correlations, suggesting they were not accurate. Five out of the eight definite glyph times that include 'H<sup>M</sup>' ('HMEPAΣ', 'of the day'; Fig. 4) conform to a model that calculates glyph times from mean lunar months—but two others certainly do not. Introducing the first lunar and solar anomalies into the analysis of the glyph times should reveal a periodic cycle of corrections following the full moon cycle (Supplementary Box 2). But the glyph times do not conform to this pattern. We conclude that the process of generation of glyph times was not sound and may remain obscure.

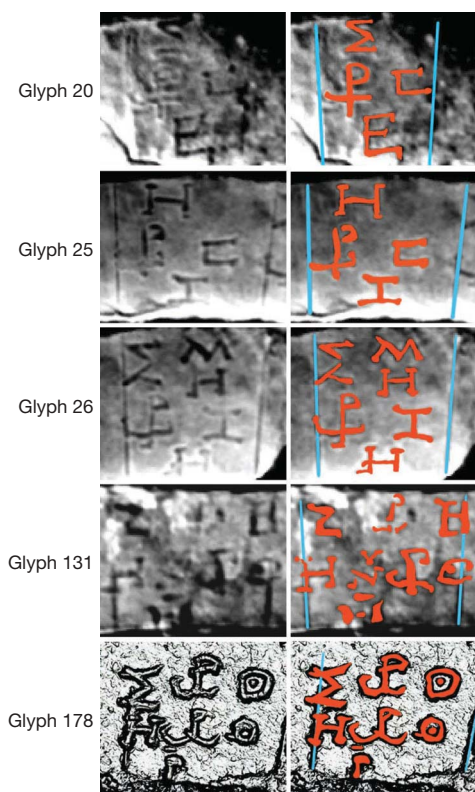
We have discovered why the Saros dial is a four-turn spiral: each quarter-turn of the dial covers a full moon cycle (Supplementary Box 2). So the apparent diameter of the Moon, which mediates the duration and type of an eclipse, is indicated by the angle of the pointer within each quarter turn of the dial.

Each Saros series eclipse occurs about 8 hours later in the day (Supplementary Box 2). After three Saros cycles (the Exeligmos) the eclipse is at nearly the same time of day. The Exeligmos dial<sup>6</sup> is divided into three sectors, with no inscription in one sector and the numbers 8 and 16 in successive sectors (Fig. 2). We conclude that these numbers tell the user how many hours to add to the glyph time to get the time of the predicted eclipse.

The newly discovered inscriptions reveal that the Antikythera Mechanism was not simply an instrument of abstract science, but exhibited astronomical phenomena in relation to Greek social institutions. It is totally unexpected that it was made for use



**Figure 3 | Deciphering the Metonic and Olympiad dials.** **a**, Representative CT slice of fragment B, showing part of the Metonic dial. The scales are 7 mm wide and the text 1.7 mm high (Supplementary Notes 3). **b**, Text in red was traced from the CT—just enough being deciphered to discover all the month names; text in blue is reconstructed (colouring in **d** follows this convention). The months here are KPANEIOΣ, ΛΑΝΟΤΡΟΠΙΟΣ, MAXANEYΣ and ΔΩΔΕΚΑΤΕΥΣ. **c**, CT slice through fragment B, showing the Olympiad dial. LA and NEMEA can be seen faintly on the left-hand side. **d**, The four sectors of the Olympiad dial are labelled LA, LB, LF and LD—years 1, 2, 3 and 4. Outside are the Panhellenic Games: year 1: ΙΣΘΜΙΑ, ΟΛΥΜΠΙΑ; year 2: NEMEA, NAA; year 3: ΙΣΘΜΙΑ, ΠΥΘΙΑ; and year 4: NEMEA, and undeciphered text. To the right of the dial are the numbers ζ (6) and ΙΑ (11) for the excluded days.



**Figure 4 | The glyphs.** A selection of the 18 known eclipse prediction glyphs (Supplementary Notes 4). Left, the raw data; X-ray CT for glyphs 20, 25, 26 and 131, and polynomial texture mapping for glyph 178. Right, the text traced in red. Most of the glyph symbols were previously decoded<sup>6</sup>. Σ, ΣΕΛΗΝΗ (Moon); Η, ΗΑΙΟΣ (Sun); Η<sup>Μ</sup>, ΗΜΕΡΑΣ (of the day); ω<sup>Ρ</sup>, ωΡΑ (hour) and the text that follows is the eclipse time in hours. Here we add Ν<sup>Υ</sup>, ΝΥΚΤΟΣ (of the night), as seen in glyph 131, and the identification of index letters at the bottom of each glyph in alphabetical order. In the consecutive glyphs 20, 25 and 26 the index letters Ε, Ζ and Η can be seen. (Ζ is always written on the Mechanism as an Ι with long serifs.) Some of the glyphs have unexplained bars over the index letter, as in glyphs 131 and 178. The index letters have profound consequences for the design of the glyph sequence.

in northwestern Greece or Sicily, rather than Rhodes as is often suggested. The Metonic calendar, the Olympiad dial and the Saros eclipse prediction scheme add new insights into the sophisticated functions of this landmark in the history of technology.

Received 28 March; accepted 2 June 2008.

- Price, D. de S. Gears from the Greeks: The Antikythera Mechanism — A calendar computer from ca. 80 BC. *Trans. Am. Phil. Soc. New Ser.*, **64**, 1–70 (1974); reprinted by Science History Publications, New York, 1975).
- Wright, M. T. Epicyclic gearing and the Antikythera Mechanism, Part I. *Antiquar. Horol.* **27**, 270–279 (2003).
- Wright, M. T. The Antikythera Mechanism: A new gearing scheme. *Bull. Sci. Instrum. Soc.* **85**, 2–7 (2005).
- Wright, M. T. Epicyclic gearing and the Antikythera Mechanism, Part II. *Antiquar. Horol.* **29**, 51–63 (2005).
- Wright, M. T. The Antikythera Mechanism and the early history of the moon-phase display. *Antiquar. Horol.* **29**, 319–329 (2006).
- Freeth, T. *et al.* Decoding the ancient Greek astronomical calculator known as the Antikythera Mechanism. *Nature* **444**, 587–591 (2006).

- Edmunds, M. *et al.* The Antikythera Mechanism research project. (<http://www.antikythera-mechanism.gr>) (2008).
- Wright, M. T. Counting months and years: The upper back dial of the Antikythera Mechanism. *Bull. Sci. Instrum. Soc.* **87**, 8–13 (2005).
- Evans, J. & Berggren, J. L. *Geminus's Introduction to the Phenomena* (Princeton Univ. Press, Princeton, 2006).
- Neugebauer, O. *A History of Ancient Mathematical Astronomy* 617 (Springer, New York, 1975).
- Toomer, G. J. *Ptolemy's Almagest* (Duckworth, London, 1984).
- Malzbender, T. & Gelb, D. Polynomial Texture Mapping (PTM). (<http://www.hpl.hp.com/research/ptm>) (2006).
- Hadland, R. *et al.* Antikythera mechanism research project: The inspection. (<http://www.xtekxray.com/applications/antikythera.html>) (2008).
- Hannah, R. *Greek & Roman Calendars: Constructions of Time in the Classical World* 170 (Duckworth, London, 2005).
- Jones, A. Calendrica, I: New Callippic dates. *Z. Papyrologie Epigraphik* **129**, 141–158 (2000).
- Trümper, C. *Untersuchungen zu den altgriechischen Monatsnamen und Monatsfolgen* (Bibliothek der Klassischen Altertumswissenschaften, NF, 2nd series, Vol. 98, Carl Winter, Heidelberg, 1997).
- Keyes, C. W. *Cicero XVI, De Re Publica*, Book 1, Sect. xiv, Para. 22 (Loeb Classical Library No. 213, Harvard Univ. Press, Cambridge, Massachusetts, 1928).
- Hultsch, F. *Pappi Alexandrini collectionis quae supersunt* 1026 (Book 8, Vol. 3, Weidmann, Berlin, 1878).
- Economou, N. A. in *Antikythera Mechanism. Astronomical Measurement Instruments from Ancient Greek Tradition* (eds Economou, N. A., Nikolantonakis, K. & Nitsiou, P.) 14 (Technology Museum of Thessaloniki, Thessaloniki, 2000).
- Dillon, M. *Pilgrims and Pilgrimage in Ancient Greece* 99–106 (Routledge, London, 1997).
- Cabanes, P. Les concours des Naia de Dodone. *Nikephoros - Zeitschrift Fur Sport und Kultur Im Altertum* **1**, 49–84 (1988).
- Esenak, F. NASA eclipse web site. (<http://eclipse.gsfc.nasa.gov/eclipse.html>) (2008).
- Britton, J. P. in *Die Rolle der Astronomie in den Kulturen Mesopotamiens* (ed. Galter, H. D.) 61–76 (rm-Druck & Verlagsgesellschaft, Graz, 1993).
- Steele, J. M. Eclipse prediction in Mesopotamia. *Arch. Hist. Exact Sci.* **54**, 421–454 (2000).
- Aaboe, A. Remarks on the theoretical treatment of eclipses in antiquity. *J. Hist. Astron.* **3**, 105–118 (1972).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was carried out under the aegis of the Antikythera Mechanism Research Project with the collaboration of M. G. Edmunds, J. Seiradakis, X. Moussas, A. Tselikas and the National Archaeological Museum in Athens. We acknowledge the essential collaboration of the Museum Director, N. Kaltsas, H. Mangou, M. Zafeiropoulou (who catalogued the fragments, revealing fragment F and several smaller fragments), G. Makris and many other contributing staff. The X-ray data were gathered by a team from X-Tek Systems (UK)/Metris (NL), led by R. Hadland, and we particularly thank A. Ramsey and A. Ray. We appreciate the support of C. Reinhart of Volume Graphics. We thank the team from Hewlett-Packard (US), led by T. Malzbender, who carried out the surface imaging. We are indebted to N. Economou (deceased), H. Kritzas, E. Georgoudakis and M. Anastasiou. We thank F. Esenak of NASA/GSFC for all the modern eclipse data and predictions. The J. F. Costopoulos Foundation partly funded T.F. and the National Bank of Greece partly funded Y.B. J.M.S. was supported by a Royal Society University Research Fellowship.

**Author Contributions** T.F. and A.J. carried out the CT analysis of the inscriptions and A.J. provided most of the interpretation and historical analysis, with input from Y.B. A.J. identified the excluded days on the Metonic dial, the Olympiad dial and the index letters in the glyphs. T.F. proposed how the Olympiad dial was turned. J.S., T.F. and A.J. contributed ideas about the glyphs and T.F. carried out the analysis of glyph sequence generation and glyph times and explained why the Saros dial has four turns. J.S. identified the role of the Exeligmos dial. All the authors contributed to the written manuscript. T.F. collated and condensed the text and designed the illustrations.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.F. ([tony@images-first.com](mailto:tony@images-first.com)).



## LETTERS

## Functionalizing hydrogen-bonded surface networks with self-assembled monolayers

Rafael Madueno<sup>1†</sup>, Minna T. Räisänen<sup>1</sup>, Christophe Silien<sup>1</sup> & Manfred Buck<sup>1</sup>

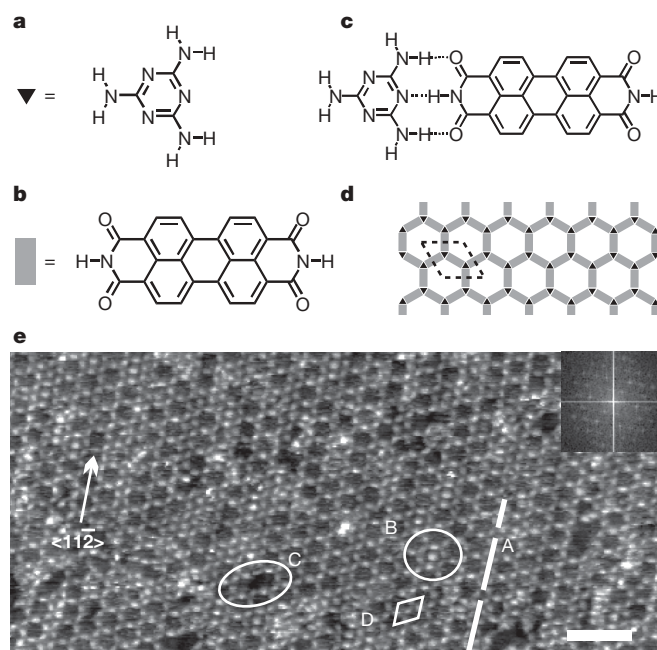
One of the central challenges in nanotechnology is the development of flexible and efficient methods for creating ordered structures with nanometre precision over an extended length scale. Supramolecular self-assembly on surfaces offers attractive features in this regard: it is a 'bottom-up' approach and thus allows the simple and rapid creation of surface assemblies<sup>1,2</sup>, which are readily tuned through the choice of molecular building blocks used and stabilized by hydrogen bonding<sup>3–8</sup>, van der Waals interactions<sup>9</sup>,  $\pi$ - $\pi$  bonding<sup>10,11</sup> or metal coordination<sup>12,13</sup> between the blocks. Assemblies in the form of two-dimensional open networks<sup>3,9,10,13–17</sup> are of particular interest for possible applications because well-defined pores can be used for the precise localization and confinement of guest entities such as molecules or clusters, which can add functionality to the supramolecular network. Another widely used method for producing surface structures involves self-assembled monolayers (SAMs)<sup>18</sup>, which have introduced unprecedented flexibility in our ability to tailor interfaces and generate patterned surfaces<sup>19–22</sup>. But SAMs are part of a top-down technology that is limited in terms of the spatial resolution that can be achieved. We therefore rationalized that a particularly powerful fabrication platform might be realized by combining non-covalent self-assembly of porous networks and SAMs, with the former providing nanometre-scale precision and the latter allowing versatile functionalization. Here we show that the two strategies can indeed be combined to create integrated network-SAM hybrid systems that are sufficiently robust for further processing. We show that the supramolecular network and the SAM can both be deposited from solution, which should enable the widespread and flexible use of this combined fabrication method.

We form our supramolecular network from 1,3,5-triazine-2,4,6-triamine (melamine; Fig. 1a) and perylene-3,4,9,10-tetracarboxylic di-imide (PTCDI; Fig. 1b), which form a synthon involving three hydrogen bonds (Fig. 1c). The three-fold symmetry of melamine and the two-fold symmetry of PTCDI give rise to a hexagonal network as shown schematically in Fig. 1d. This bimolecular network is particularly flexible because pore size and shape can be varied by using analogues of PTCDI, and functionality can be added by attaching side groups to the aromatic rings.

So far, the synthesis and further modification of the PTCDI-melamine network have always been performed in an ultra-high-vacuum environment<sup>3,23</sup>, which restricts the assembly to molecules that can be sublimed and makes further processing of the network difficult. A solution-based fabrication strategy avoids these limitations and might even allow subsequent processing using SAMs of thiols and related compounds, which offer a plethora of possibilities for surface modification<sup>18,22</sup>.

We successfully accomplished a solution-based assembly of the network through adsorption on gold, by using a mixture of PTCDI

and melamine in dimethylformamide. The scanning tunnelling microscope (STM) image of the resultant network in Fig. 1e reveals the honeycomb arrangement of the PTCDI molecules, which are the moieties resolved on this scale. The period of the honeycomb is 35 Å, which corresponds to a  $(7\sqrt{3} \times 7\sqrt{3})R30^\circ$  unit cell<sup>23</sup>. In contrast to the 50% coverage observed in an earlier ultra-high-vacuum experiment<sup>3</sup>, we find that the network forms over extended areas. The network structure overall is very regular and there are no major discontinuities, but some imperfections are discernible. The first, highlighted by the dashed line A in Fig. 1e, is a fault line with neighbouring hexagons meeting at a vertex instead of sharing an edge. The second is an additional PTCDI molecule trapped in a pore (marked by ellipse B). A third imperfection is a missing PTCDI molecule (ellipse C in Fig. 1e), thus joining two adjacent cells.



**Figure 1 | Supramolecular network of melamine-PTCDI self-assembled on Au(111).** **a–c**, Structures of melamine (**a**) and PTCDI (**b**) and the bonding motif (**c**). **d**, Schematic diagram of the network with the unit cell indicated by a dotted rhombus. **e**, STM image of network recorded in ambient conditions. The dashed line A highlights a fault line. Circled areas B and C mark a pore hosting a PTCDI molecule and a missing PTCDI molecule, respectively. The  $(7\sqrt{3} \times 7\sqrt{3})R30^\circ$  unit cell (D) corresponding to a 35-Å period of the honeycomb is also indicated. The inset shows a Fourier transform. Scale bar, 10 nm.

<sup>1</sup>EaStCHEM School of Chemistry, University of St Andrews, North Haugh, St Andrews KY16 9ST, UK. <sup>†</sup>Present address: Departamento de Química Física, Universidad de Córdoba, Campus de Rabanales, 14014 Córdoba, Spain.



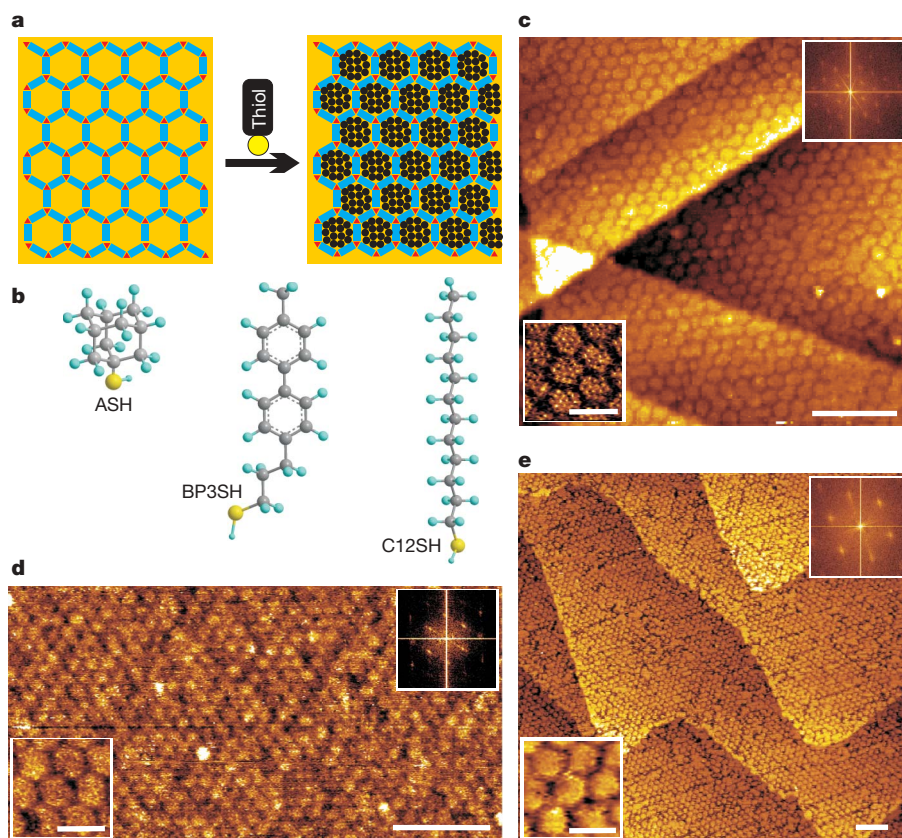
The solution-based preparation makes the network a readily available template, but the scope for further modification and use depend on its stability under the conditions of subsequent processing. As illustrated in Fig. 2a, we aim for further modification with SAMs. A precise estimation of network stability under relevant conditions is not possible because of the lack of precise data for the network, in particular for the adsorption energies of PTCDI and melamine. However, we can use the hydrogen-bond energy per synthon (values range from  $70 \text{ kJ mol}^{-1}$  (ref. 24) to  $90 \text{ kJ mol}^{-1}$  (ref. 25)) to calculate total network binding energies of  $140\text{--}180 \text{ kJ mol}^{-1}$  and  $210\text{--}270 \text{ kJ mol}^{-1}$  per PTCDI and melamine molecule, respectively. The adsorption energies of PTCDI and melamine are taken to be similar to those of other aromatic hydrocarbons<sup>26,27</sup>, which range from  $50$  to  $200 \text{ kJ mol}^{-1}$ . With this approach, we estimate the binding energy of a network molecule to fall in the range  $200\text{--}470 \text{ kJ mol}^{-1}$ , which is higher than the  $160\text{--}200 \text{ kJ mol}^{-1}$  of an Au–S bond<sup>18,22</sup>. However, considering that more than one thiol molecule can be adsorbed in the area occupied by PTCDI and melamine, we conclude that thiol adsorption can match the network energetically.

To investigate to what extent thiols can be adsorbed in the network we chose three types of molecule (see Fig. 2b) differing in the stability of the respective SAMs. One is small and rigid and has rather weak intermolecular interactions (adamantane thiol; ASH)<sup>28</sup>; the other two show more pronounced intermolecular interactions, one consisting of a rigid aromatic moiety combined with an aliphatic spacer ( $\omega$ -(4'-methylbiphenyl-4-yl)propane thiol; BP3SH) and the other of a flexible alkane chain (dodecane thiol; C12SH). Large-scale STM images of the resultant structures (Fig. 2c–e) show that the network acts as template for all three types of molecule, with high-resolution images and Fourier transforms (see insets) confirming that in all cases the hexagonal pattern is well maintained after thiol adsorption.

In contrast to the empty network, in which the molecules appear as protrusions (Fig. 1), filling the network pores inverts the height contrast so that the presence of the network is reflected by the appearance of hexagonal grooves. It is worth noting that as a result of the rigidity of adamantane thiol it was even possible to achieve molecular resolution (Fig. 2c, inset).

Figure 2 demonstrates that the supramolecular network serves as a general template for a range of thiol molecules that form SAMs differing substantially in structure, intermolecular interactions and stability. However, we note that the details of the preparation protocol are dependent on the SAM molecule used, and reflect the above estimated similarity of SAMs and network with regard to their energetics. For adamantane thiol, which is known to form SAMs that are not very stable (in comparison with SAMs formed from alkane thiols, for example), immersion time is not critical: the pores of the network are filled within seconds, and the network itself is perfectly stable against displacement by ASH. In contrast, for the other two molecules, prolonged exposure of the network to a solution of the respective thiol molecules ultimately results in the displacement of the network and the formation of a uniform SAM. However, there is a pronounced difference between the rate at which the pores are filled and the rate at which the network is displaced, so that selective adsorption in the pores while maintaining the network structure can be controlled kinetically, as demonstrated by Fig. 2.

Once formed, the SAM–network hybrid structure is stable in a liquid environment and can be processed further, as we illustrate here with the electrochemical deposition of Cu in the underpotential deposition (UPD) region. The experiment, sketched in Fig. 3a, involves the mounting of a sample with the SAM–network hybrid in an electrochemical cell containing  $\text{Cu}^{2+}$  ions. A potential in the UPD region of Cu (that is, positive of the Nernst potential) is then



**Figure 2 | Generation of a network–SAM hybrid structure.** **a**, Scheme of filling the cells of the PTCDI–melamine network by thiols. **b**, Structures of the thiols studied. **c–e**, STM images of the hybrid structures on Au(111)/mica: network filled with ASH (**c**), C12SH (**d**) and BP3SH (**e**). The insets at

the lower left and upper right corners of the STM images show high-resolution images and Fourier transforms, respectively. Scale bars, 20 nm (large-scale images) and 5 nm (insets).

applied, which causes the insertion of a monolayer of Cu between the Au substrate and the thiol molecules<sup>29</sup>. The Cu insertion renders the thiol–substrate bond more stable and could be used for further patterning<sup>30</sup>. After deposition, the sample was removed from the cell and investigated by STM in ambient environment, with the image (Fig. 3b) revealing that the pattern of the hybrid structure is preserved.

To probe the insertion of Cu, experiments were performed with a deposition time chosen such that Cu UPD had not yet occurred homogeneously across the whole sample. In the STM image (Fig. 3c), the hexagonal structure is discernible in both the unaltered and the UPD areas. In contrast, the corresponding height profile (Fig. 3d) reveals an increase in height *S* as a result of Cu UPD. A particular feature of Fig. 3d is the difference in the corrugation between the UPD and the unaltered area, respectively. On the UPD part the corrugation *A* is significantly larger than the corrugation *B* of the unaltered area. This strongly suggests that Cu is inserted only between thiol and substrate and not between network and substrate as illustrated in Fig. 3a; that is, the network acts as a diffusion barrier. This interpretation is corroborated by the appearance of isolated UPD islands (marked by arrows in Fig. 3c) in which only one cell is filled. The suppression of Cu diffusion at the interface by the network makes the hybrid system very different from a uniform SAM in which Cu UPD cannot be confined because of the lack of such a diffusion barrier<sup>29</sup>. We also note that, in comparison with densely packed SAMs, intercalation of the Cu ions at the thiol/substrate interface is greatly facilitated and faster for the hybrid system as a result of the more open structure. Overall, the hybrid system renders UPD on the nanometre scale much more controllable than when using a SAM without network.

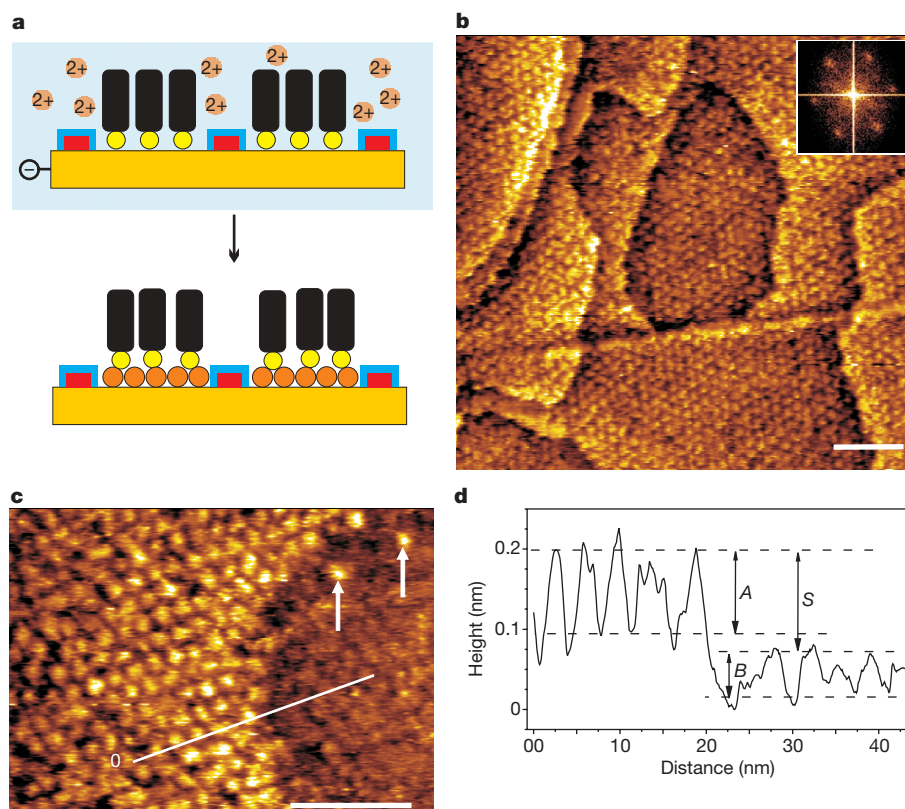
The results presented here show that the honeycomb network involving three hydrogen bonds per synthon demonstrates sufficient

stability to act as template in subsequent processes. The combination of the network with SAMs offers considerable design flexibility, with the network providing an exact definition of structures in the substrate plane and the SAM permitting separate surface modification. Whether the SAM-modified pores serve as active sites for a precise localization of species through chemical interactions or whether they are used as blocking sites to direct species to the network molecules, the hybrid system can provide control on a length scale and at a precision not readily achievable otherwise. We believe that this feature, and the fact that the hybrid system is accessible through exclusively solution-based processing, will facilitate a wide range of fundamental studies into how confined nanometre-sized geometries can influence phenomena as diverse as electrochemistry, tribology and wetting.

## METHODS SUMMARY

PTCDI (at least 98% pure; Alfa Aesar) and melamine (99.9% pure; Sigma-Aldrich) were used without further purification. The mixture of PTCDI and melamine used for the experiments was prepared from saturated solutions of PTCDI and melamine in dimethylformamide that were diluted, typically 1:25 and 1:4, respectively. Au/mica substrates (300 nm gold; Georg Albert PVD) were flame-annealed before immersion in the PTCDI/melamine solution. Immersion times for network formation were up to 3 min at temperatures between 325 and 400 K, with 1 min and 371 K as a typical combination of parameters. After removal from solution, samples were blown dry in a stream of nitrogen or argon. For thiol adsorption experiments, network/Au/mica samples were immersed in a 1 mM solution of the respective thiol (ASH (99.9% pure; Sigma-Aldrich), C12SH (at least 98% pure; Sigma-Aldrich) or BP3SH (for synthesis see ref. 43 in ref. 29)) in ethanol at 18–21 °C. Immersion times were varied between 3 s and 24 h. After immersion, samples were thoroughly rinsed with ethanol and blown dry with nitrogen.

Partial Cu UPD was achieved in 50 mM CuSO<sub>4</sub>, 0.5 μM H<sub>2</sub>SO<sub>4</sub> (aqueous) by setting the sample potential at +100 mV against Cu/Cu<sup>2+</sup> for 10 s in a polytetrafluoroethylene electrochemical cell. The sample was then rinsed with



**Figure 3 | UPD of Cu on Au(111) modified by an adamantane-thiol-filled PTCDI-melamine network.** **a**, Illustration of electrochemical Cu deposition in pores of the network at the thiol/Au interface. **b**, **c**, STM images of samples taken in ambient atmosphere after Cu UPD: complete UPD (**b**) and partial

UPD (**c**). Scale bars, 20 nm. Arrows in **c** mark isolated cells of Cu UPD. **d**, Height profile along the line given in **c**, with the origin marked by '0'. Corrugations are *A* = 1.15 Å on UPD areas and *B* = 0.5 Å on unaltered areas. The height difference between UPD and unaltered areas is *S* = 1.3 Å.

deionized water and blown dry with nitrogen. Complete Cu UPD coverage was achieved by repeating the same procedure once.

Samples were characterized under ambient by STM with a PicoPlus STM (Molecular Imaging). Bias and currents were typically in the range 250–800 mV (tip positive) and 5–80 pA. Figure 1e is despeckled, all other images are presented as acquired.

Received 3 December 2007; accepted 9 May 2008.

- De Feyter, S. & De Schryver, F. C. Two-dimensional supramolecular self-assembly probed by scanning tunneling microscopy. *Chem. Soc. Rev.* **32**, 139–150 (2003).
- Barth, J. V. Molecular architectonic on metal surfaces. *Annu. Rev. Phys. Chem.* **58**, 375–407 (2007).
- Theobald, J. A., Oxtoby, N. S., Phillips, M. A., Champness, N. R. & Beton, P. H. Controlling molecular deposition and layer structure with supramolecular surface assemblies. *Nature* **424**, 1029–1031 (2003).
- Li, Z., Han, B., Wan, L. J. & Wandlowski, T. Supramolecular nanostructures of 1,3,5-benzene-tricarboxylic acid at electrified Au(111)/0.05 M H<sub>2</sub>SO<sub>4</sub> interfaces: An *in situ* scanning tunneling microscopy study. *Langmuir* **21**, 6915–6928 (2005).
- Pinheiro, L. S. & Temperini, M. L. A. Pyridine and pyridine carboxylic acids as guests in a bidimensional hydrogen bond structure analyzed by scanning tunneling microscopy. *Surf. Sci.* **601**, 1836–1843 (2007).
- Canas-Ventura, M. E. *et al.* Self-assembly of periodic bicomponent wires and ribbons. *Angew. Chem. Int. Ed.* **46**, 1814–1818 (2007).
- Payer, D. *et al.* Ionic hydrogen bonds controlling two-dimensional supramolecular systems at a metal surface. *Chem. Eur. J.* **13**, 3900–3906 (2007).
- Kampschulte, L., Griessl, S., Heckl, W. M. & Lackinger, M. Mediated coadsorption at the liquid–solid interface: Stabilization through hydrogen bonds. *J. Phys. Chem. B* **109**, 14074–14078 (2005).
- Furukawa, S. *et al.* Structural transformation of a two-dimensional molecular network in response to selective guest inclusion. *Angew. Chem. Int. Ed.* **46**, 2831–2834 (2007).
- Mena-Osteritz, E. & Bäuerle, P. Complexation of C<sub>60</sub> on a cyclothiophene monolayer template. *Adv. Mater.* **18**, 447–451 (2006).
- Schenning, A. & Meijer, E. W. Supramolecular electronics; nanowires from self-assembled  $\pi$ -conjugated systems. *Chem. Commun.* 3245–3258 (2005).
- Diaz, D. J., Bernhard, S., Storrier, G. D. & Abruna, H. D. Redox active ordered arrays via metal initiated self-assembly of terpyridine based ligands. *J. Phys. Chem. B* **105**, 8746–8754 (2001).
- Stepanow, S. *et al.* Steering molecular organization and host–guest interactions using two-dimensional nanoporous coordination systems. *Nature Mater.* **3**, 229–233 (2004).
- Stöhr, M., Wahl, M., Spillmann, H., Gade, L. H. & Jung, T. A. Lateral manipulation for the positioning of molecular guests within the confinements of a highly stable self-assembled organic surface network. *Small* **3**, 1336–1340 (2007).
- Spillmann, H. *et al.* A two-dimensional porphyrin-based porous network featuring communicating cavities for the templated complexation of fullerenes. *Adv. Mater.* **18**, 275–279 (2006).
- Lu, J. *et al.* Template-induced inclusion structures with copper(II) phthalocyanine and coronene as guests in two-dimensional hydrogen-bonded host networks. *J. Phys. Chem. B* **108**, 5161–5165 (2004).
- Stepanow, S. *et al.* Surface-assisted assembly of 2D metal-organic networks that exhibit unusual threefold coordination symmetry. *Angew. Chem. Int. Ed.* **46**, 710–713 (2007).
- Schreiber, F. Self-assembled monolayers: from ‘simple’ model systems to biofunctionalized interfaces. *J. Phys. Condens. Matter* **16**, R881–R900 (2004).
- Gooding, J. J., Mearns, F., Yang, W. R. & Liu, J. Q. Self-assembled monolayers into the 21<sup>st</sup> century: Recent advances and applications. *Electroanal.* **15**, 81–96 (2003).
- Mrksich, M. A surface chemistry approach to studying cell adhesion. *Chem. Soc. Rev.* **29**, 267–273 (2000).
- Thom, I., Hähner, G. & Buck, M. Replicative generation of metal microstructures by template directed electrometallization. *Appl. Phys. Lett.* **87**, 024101 (2005).
- Love, J. C., Estroff, L. A., Kriebel, J. K., Nuzzo, R. G. & Whitesides, G. M. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem. Rev.* **105**, 1103–1170 (2005).
- Perdigao, L. M. A. *et al.* Bimolecular networks and supramolecular traps on Au(111). *J. Phys. Chem. B* **110**, 12539–12542 (2006).
- Weber, U. K. *et al.* Role of interaction anisotropy in the formation and stability of molecular templates. *Phys. Rev. Lett.* **100**, 156101 (2008).
- Aakeroy, C. B. & Seddon, K. R. The hydrogen-bond and crystal engineering. *Chem. Soc. Rev.* **22**, 397–407 (1993).
- Baldacchini, C., Mariani, C. & Betti, M. G. Adsorption of pentacene on filled d-band metal surfaces: Long-range ordering and adsorption energy. *J. Chem. Phys.* **124**, 154702 (2006).
- Bilic, A., Reimers, J. R., Hush, N. S., Hoft, R. C. & Ford, M. J. Adsorption of benzene on copper, silver, and gold surfaces. *J. Chem. Theory Comput.* **2**, 1093–1105 (2006).
- Dameron, A. A., Charles, L. F. & Weiss, P. S. Structures and displacement of 1-adamantanethiol self-assembled monolayers on Au{111}. *J. Am. Chem. Soc.* **127**, 8697–8704 (2005).
- Silien, C. & Buck, M. On the role of extrinsic and intrinsic defects in the underpotential deposition of Cu on thiol-modified Au(111) electrodes. *J. Phys. Chem. C* **112**, 3881–3890 (2008).
- Oyamatsu, D., Kanemoto, H., Kuwabata, S. & Yoneyama, H. Nanopore preparation in self-assembled monolayers of alkanethiols with use of the selective desorption technique assisted by underpotential deposition of silver and copper. *J. Electroanal. Chem.* **497**, 97–105 (2001).

**Acknowledgements** We are grateful to J. D. E. T. Wilton-Ely for his contribution to the synthesis of BP3SH. This work was supported by the UK Engineering Physical Sciences Research Council (EPSRC). M.R. acknowledges support from the Academy of Finland.

**Author Contributions** All authors contributed to the design of the experiments. The preparation and characterization of networks and hybrid systems were performed by R.M. and M.T.R.. Experiments and analysis related to electrochemistry were conducted by C.S. All authors contributed to the manuscript, with M.B. and R.M. leading.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.B. (mb45@st-and.ac.uk) or R.M. (rafael.madueno@uco.es).



## LETTERS

# Primary carbonatite melt from deeply subducted oceanic crust

M. J. Walter<sup>1</sup>, G. P. Bulanova<sup>1</sup>, L. S. Armstrong<sup>1</sup>, S. Keshav<sup>2</sup>, J. D. Blundy<sup>1</sup>, G. Gudfinnsson<sup>2</sup>, O. T. Lord<sup>1</sup>, A. R. Lennie<sup>3</sup>, S. M. Clark<sup>4</sup>, C. B. Smith<sup>5</sup> & L. Gobbo<sup>6</sup>

Partial melting in the Earth's mantle plays an important part in generating the geochemical and isotopic diversity observed in volcanic rocks at the surface<sup>1</sup>. Identifying the composition of these primary melts in the mantle is crucial for establishing links between mantle geochemical 'reservoirs' and fundamental geodynamic processes<sup>2</sup>. Mineral inclusions in natural diamonds have provided a unique window into such deep mantle processes<sup>3–8</sup>. Here we provide experimental and geochemical evidence that silicate mineral inclusions in diamonds from Juina, Brazil, crystallized from primary and evolved carbonatite melts in the mantle transition zone and deep upper mantle. The incompatible trace element abundances calculated for a melt coexisting with a calcium-titanium-silicate perovskite inclusion indicate deep melting of carbonated oceanic crust, probably at transition-zone depths. Further to perovskite, calcic-majorite garnet inclusions record crystallization in the deep upper mantle from an evolved melt that closely resembles estimates of primitive carbonatite on the basis of volcanic rocks. Small-degree melts of subducted crust can be viewed as agents of chemical mass-transfer in the upper mantle and transition zone, leaving a chemical imprint of ocean crust that can possibly endure for billions of years.

The Juina region is well known as a source of 'ultra-deep' mineral inclusions in diamonds<sup>5,6,9,10</sup>. Here we report on mineral inclusions exposed by polishing three nitrogen-free type II diamonds (J1, J9 and J10) from the Collier 4 kimberlite pipe. Composite grains of calcium silicate (CaSiO<sub>3</sub>) plus calcium titanate (CaTiO<sub>3</sub>) are found in two diamonds, J1 and J10, and three inclusions of calcic-majoritic garnet occur in diamonds J1 and J9. Optical and cathodoluminescent observations are consistent with syngensis of inclusions and host diamonds<sup>11</sup>.

As shown in Fig. 1a, two chemically distinct regions in the composite grains have nearly end-member compositions in the system CaTiO<sub>3</sub>–CaSiO<sub>3</sub> (Supplementary Table 1), indicative of unmixing from originally homogeneous minerals. Although rare, other composite inclusions of CaTiO<sub>3</sub> and CaSiO<sub>3</sub> have recently been reported in Juina detrital diamonds<sup>6,9,10,12</sup>. Previous workers have interpreted these grains as retrograde exsolution products from pre-existing, single-phase Ca(Ti,Si)O<sub>3</sub>-perovskite solid solutions originating at transition zone or lower mantle depths, and their compositions were used as evidence of deep subduction of a Ca-rich lithology<sup>9,12</sup>.

High-pressure phase relations in the system CaTiO<sub>3</sub>–CaSiO<sub>3</sub> show immiscibility at pressures of less than about 9 GPa at 1,200 °C (Fig. 1a)<sup>13</sup>. We interpret the composite inclusions in J9 and J10 as retrograde perovskite solid solutions that unmixed on diamond exhumation from deeper in the mantle to final equilibration pressures of 6 to 7 GPa (~250 km), as deduced from the CaSiO<sub>3</sub> content of the CaTiO<sub>3</sub>-rich phase. Figure 1 shows that a complete perovskite

solid solution exists for the estimated inclusion compositions at pressures of greater than ~9 to 11 GPa (1,200 °C), indicating that their primary depth of origin is a minimum of about 300 km.

The garnets in diamonds J1 and J9 have minor majorite components but are conspicuous in their very high CaO contents (~8–15 wt%), near absence of Cr<sub>2</sub>O<sub>3</sub> (<0.05%), and high Na<sub>2</sub>O and TiO<sub>2</sub> (Supplementary Table 1), all features typically interpreted as 'eclogitic', genetically linking the minerals to subducted oceanic crust<sup>7</sup>. On the basis of the composition of the garnet near the rim of J1, an equilibrium pressure of about 7 GPa is indicated (Fig. 1b), consistent with the pressure deduced for the centrally located perovskite inclusion.

Further constraints on the petrogenesis and primary depth of origin of the perovskite inclusions come from high-pressure phase relations as depicted in the system CaSiO<sub>3</sub>–CaTiO<sub>3</sub>–MgSiO<sub>3</sub>. The perovskite inclusions show a near absence of MgSiO<sub>3</sub> component (≤0.2 mol%), a feature noted previously for Ti-poor CaSiO<sub>3</sub>-perovskite inclusions in diamond<sup>5,14</sup>. At transition zone and lower mantle depths, Ca-rich perovskite in mantle peridotite and eclogite is saturated in MgSiO<sub>3</sub> component because it coexists with either majorite-rich garnet or Mg-rich perovskite. Inclusions that represent entrapped mantle minerals should also show MgSiO<sub>3</sub> saturation. Figure 2a shows high-pressure experimental data from the literature for the compositions of Ca-perovskite coexisting with majorite or Mg-perovskite in natural peridotite and eclogite bulk compositions<sup>15–17</sup> as projected onto the CaSiO<sub>3</sub>–CaTiO<sub>3</sub>–MgSiO<sub>3</sub> plane. At the experimental temperatures (1,200–2,430 °C), Ca-perovskite has about 3 to 5 mol% MgSiO<sub>3</sub> component, which is much more than observed in diamond-hosted inclusions. Furthermore, the CaTiO<sub>3</sub>-component of Ca-rich perovskites, even in more Ti-rich eclogitic compositions, is much lower than in the inclusions in J1 and J10.

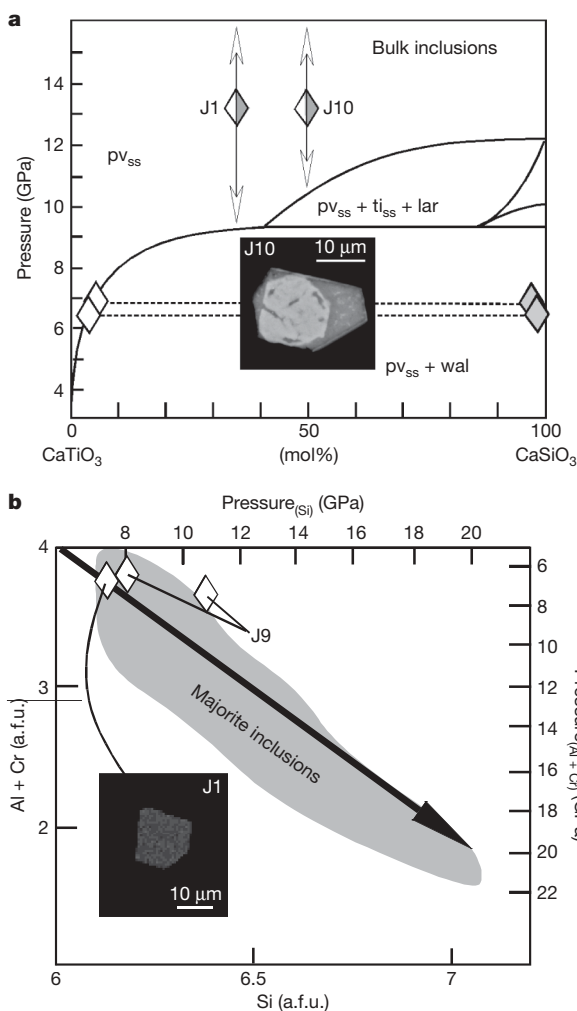
To determine the effects of titanium on MgSiO<sub>3</sub> solubility in Ca(Ti,Si)O<sub>3</sub>-perovskite solid solutions saturated in Mg-perovskite, we made high pressure experiments in the system CaSiO<sub>3</sub>–CaTiO<sub>3</sub>–MgSiO<sub>3</sub> and the results are shown in Fig. 2b. We found that at all pressures investigated (from 20 to 50 GPa, 1,227–2,227 °C) the MgSiO<sub>3</sub> solubility in Ca-perovskite increased substantially with Ti content. Phase relations show that Ti-rich Ca(Ti,Si)O<sub>3</sub>-perovskites have an abundant MgSiO<sub>3</sub> component when coexisting with Mg-perovskite. The near absence of MgSiO<sub>3</sub> component in perovskite inclusions in J1 and J10 effectively precludes an origin in Ti-rich, MgSiO<sub>3</sub>-saturated lithologies at lower mantle conditions. Thus, our observations to this point constrain the depth of origin of the perovskite inclusions to between about 300 km and the onset of Mg-perovskite stability (~670–700 km), and also apparently make a sub-solidus origin in either peridotite or eclogite highly improbable.

<sup>1</sup>Department of Earth Sciences, University of Bristol, Queen's Road, Bristol BS8 1RJ, UK. <sup>2</sup>Bayerisches Geoinstitut, Universität Bayreuth, D-95440 Bayreuth, Germany. <sup>3</sup>Daresbury Laboratory, Keckwick Lane, Warrington WA4 4AD, UK. <sup>4</sup>Advanced Light Source, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. <sup>5</sup>Rio Tinto Mining and Exploration Ltd., Paddington, London W2 6LG, UK. <sup>6</sup>Rio Tinto Desinvestimentos Minerais Ltda, SIA Trecho 2, Lote 720, Brasília, DF 71.200-020, Brazil.

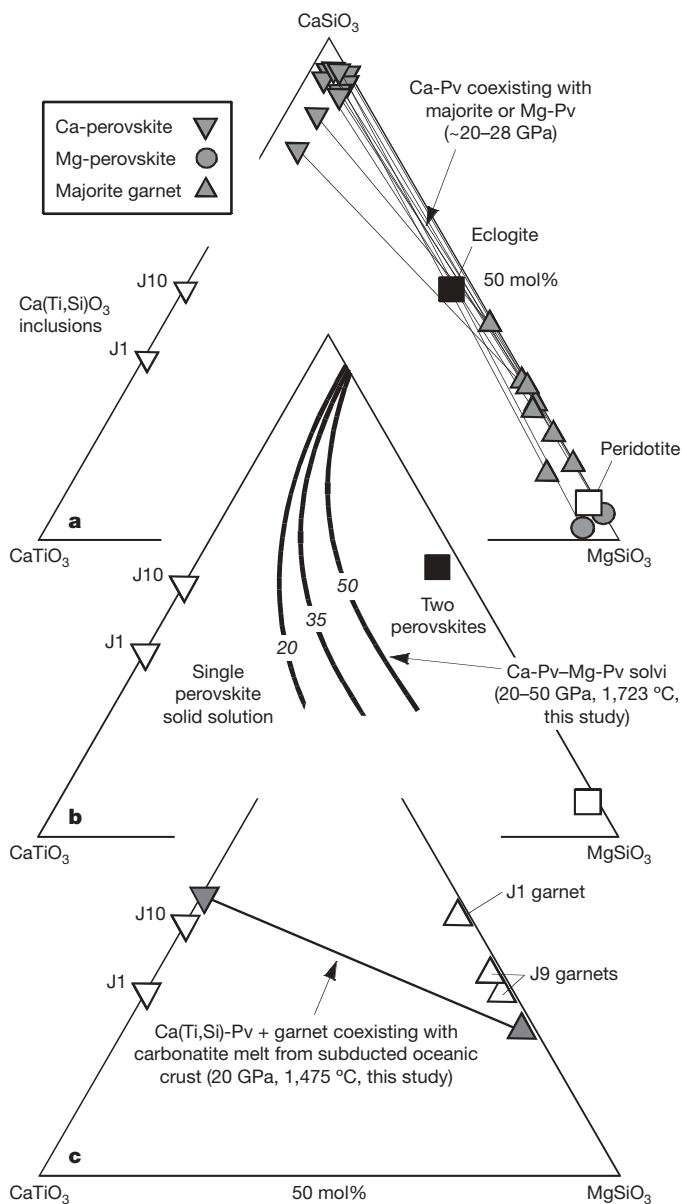
Further evidence for the petrogenesis of the mineral inclusions comes from their trace element chemistry. We analysed perovskite and garnet inclusions in diamonds J1 and J9 for trace elements using a secondary ion microprobe at the University of Edinburgh. Data are presented on relative compatibility 'spidergrams' in Fig. 3a, and show extremely elevated abundances of many 'magmaophile' trace elements when normalized to primitive mantle<sup>18</sup>. The extreme enrichment in elements such as thorium, niobium and the rare earth elements (REE) is hard to reconcile with subsolidus paragenesis in lithologies such as peridotite, harzburgite or eclogite. For example, caesium is present in the J1 CaTiSi-perovskite at a level of  $\sim 10^4$  relative to primitive mantle. Even if this inclusion represents only 0.1% of a mantle lithology, it would still yield a bulk composition enriched by at least 10 times primitive mantle. In contrast, extreme enrichment in incompatible elements is consistent with, and often signatory of, the involvement of a low-degree melt. Considering the inclusions are hosted in diamond, the melt must also be sufficiently carbon-rich to

saturate in diamond. Previous experiments confirmed that reduction of carbonate-rich melt is an effective means of diamond synthesis in mantle silicates<sup>19,20</sup>.

To test further our hypothesis that the mineral inclusions crystallized from carbon-rich low-degree melts, we performed melting experiments in the system  $\text{CaO-MgO-Al}_2\text{O}_3\text{-SiO}_2\text{-TiO}_2\text{-CO}_2$ . Because of the eclogitic characteristics of the inclusions, we chose a



**Figure 1 | Mineral chemistry and geobarometry of perovskite and garnet mineral inclusions in Juina diamonds.** **a**, Phase relations in the system  $\text{CaTiO}_3\text{-CaSiO}_3$  at 1,200 °C (ref. 13). Shown on the diagram are the compositions of the  $\text{CaSiO}_3$  and  $\text{CaTiO}_3$  regions of the composite inclusions in diamonds J1 and J10. Estimates of the bulk inclusion compositions on the basis of area proportion are also shown. Inset, back-scattered electron image of inclusion J10. On the basis of the  $\text{CaSiO}_3$  component of the  $\text{CaTiO}_3$  phase, exsolution pressures of 6 to 7 GPa are indicated. lar, larnite; pv, perovskite; ss, solid solution; ti, titanite; wal, walstromite. **b**, Systematics of (Al + Cr) versus Si per atomic formula unit (a.f.u.) in garnet inclusions from diamonds J1 and J9. Calibration of this relationship as a barometer, giving pressure shown as  $\text{Pressure}_{(\text{Si})}$ , indicates equilibration pressures of 7 (J1) to 11 (J9) GPa for these inclusions<sup>7</sup>. Scale bars, 10  $\mu\text{m}$ .

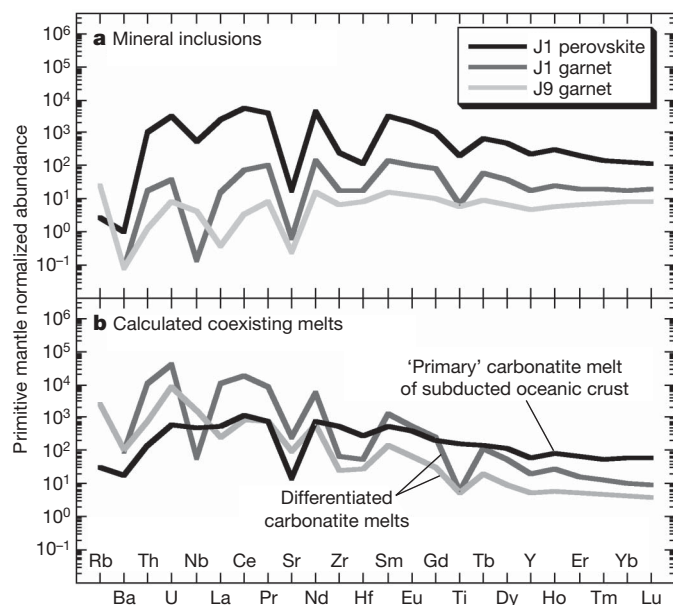


**Figure 2 | Subsolvus and melting phase relations in the compositional join  $\text{CaTiO}_3\text{-CaSiO}_3\text{-MgSiO}_3$ .** The estimated compositions of the composite inclusions in diamonds J1 and J10 are shown as inverted white triangles. Projected bulk compositions of natural eclogite and peridotite are shown as black and white squares, respectively. **a**, Ca-perovskite (Ca-Pv) coexisting with majorite or Mg-perovskite (Mg-Pv) as determined in previous experiments on natural peridotite and eclogite bulk compositions at 20 to 28 GPa and 1,200–2,430 °C (refs 15–17). Ca-rich perovskites are shown as inverted dark triangles, majorite as dark triangles, and Mg-perovskites as dark circles. **b**, The boundary curves show the solubility of the  $\text{MgSiO}_3$  component in  $\text{Ca}(\text{Si,Ti})\text{O}_3$ -perovskite solid solution coexisting with Mg-perovskite (CaPv-MgPv) at 20–50 GPa and 1,723 °C as determined in this study. **c**, The compositions of  $\text{Ca}(\text{Si,Ti})\text{O}_3$ -perovskite solid solution (inverted dark triangle) that coexist with calcic majorite (dark triangle) and Ca-rich carbonatite melt from model carbonated eclogite at 20 GPa, 1,475 °C as determined in this study. Garnet inclusions from diamonds J1 and J9 are shown as white triangles.

model oceanic crust composition. Figure 2c shows the compositions of  $\text{Ca}(\text{Ti,Si})\text{O}_3$ -perovskite and majorite coexisting with carbonated melt at 20 GPa and 1,475 °C. We find that carbonatite melt from model eclogite is  $\text{TiO}_2$  and  $\text{CaO}$  rich [ $(\text{Ca}/(\text{Ca} + \text{Mg}))_{\text{molar}} \approx 0.7$ ]. The coexisting  $\text{Ca}(\text{Ti,Si})\text{O}_3$ -perovskite has a high Ti content and a very low  $\text{MgSiO}_3$  content that is consistent with the compositions of the perovskite inclusions in J1 and J10. The coexisting majorite garnet is very calcic<sup>21</sup>, with a composition akin to the calcic garnet inclusions. Thus, the major-element mineral chemistry and the elevated trace element abundances are consistent with crystallization of the inclusions from carbonatitic melts derived from eclogite in the mantle transition zone.

We quantitatively calculated the trace element (*i*) concentrations of melts that could coexist with the mineral inclusions using the relation  $C_i^{\text{melt}} = C_i^{\text{solid}}/D$ , where the solid concentrations,  $C_i^{\text{solid}}$ , and mineral melt partition coefficients,  $D$ , are known<sup>21,22</sup> (Supplementary Tables 1 and 2). The melt calculated to coexist with  $\text{CaTiSi}$ -perovskite, or 'perovskite melt', is shown in Fig. 3b and has enriched but relatively unfractionated REE (lanthanum/lutetium  $\sim 5$ ), is not depleted in the high field strength elements (HFSE) titanium, hafnium, zirconium and niobium, but is strongly depleted in the large ion lithophile elements (LILE) strontium, barium and rubidium relative to other elements of similar incompatibility.

The characteristics of the perovskite melt are qualitatively explicable in terms of the relative abundances of elements expected in deeply subducted oceanic crust. During subduction and slab dehydration in the upper mantle the LILE can be highly soluble in the fluids and melts expelled from oceanic crust, whereas REE and HFSE may be retained in minerals such as garnet (heavy REE), allanite (light REE) and rutile (HFSE)<sup>23–25</sup>. On the basis of the characteristic trace element concentrations—together with the major-element constraints provided above—we propose that the  $\text{CaTiSi}$ -perovskite inclusion in J1 crystallized from a low-degree, primary carbonatite melt derived from deeply subducted oceanic crust.

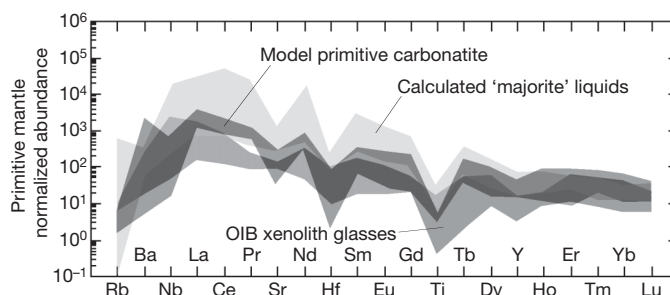


**Figure 3 | Relative compatibility diagrams showing trace element abundances of mineral inclusions and calculated coexisting melts.** Data are normalized to primitive mantle<sup>18</sup>. **a**, Element abundances measured by SIMS in  $\text{Ca}(\text{Ti,Si})\text{O}_3$  inclusions in diamond J1 and  $\text{Ca}$ -rich majorite garnets in J1 and J9. **b**, Calculated melts that could coexist with the mineral inclusions. Melt compositions were calculated using experimental mineral/melt partition coefficients from the literature appropriate to the pressure–temperature conditions of the lowermost upper mantle and transition zone (see Supplementary Table 2).

The melts calculated to coexist with majoritic garnets ('majorite melt') are also highly enriched but show several distinguishing features (Fig. 3b). The REE are much more fractionated (lanthanum/lutetium  $\sim 10^2$  to  $10^3$ ) and there are large relative depletions in HFSE, especially for the J9 garnet. We suggest that the majorite melt is either evolved from a primary melt by fractionation of high-pressure phases such as majorite and ilmenite, or perhaps it is derived from a secondary melt of mantle peridotite previously metasomatized by carbonatitic melts from eclogite. For example, a primary calcic-carbonatite melt may form in the transition zone and crystallize liquidus minerals including diamond and  $\text{CaTiSi}$ -perovskite in an eclogite body. Migration of such melt into peridotite should cause dissolution of magnesium silicate minerals and solidification. Later upwelling of this metasomatized mantle, perhaps related to plume upwelling or flow of material around a deepening cratonic keel, may cause metasomatized mantle to melt forming a new more  $\text{Mg}$ -rich carbonatite from which diamond and majorite crystallize.

Figure 4 shows the trace-element abundances of melts calculated to coexist with calcic-majorite garnet inclusions from this study and the literature. These are compared with literature estimates of primitive carbonatite magmas on the basis of volcanic rocks, and to carbonatitic glasses in xenoliths from ocean island basalts<sup>26–31</sup>. The fields show similarities both in absolute concentrations and in patterns, suggesting that the calculated majorite melts exemplify a pervasive metasomatizing component in the deep upper mantle, which is identifiable in xenolith glasses, xenolith mineral chemistry, ocean island basalt magma chemistry and diamond inclusions<sup>21,26–29,32</sup>. We suggest that this widely dispersed and chemically distinct signature is imparted by low-degree carbonatite melt derived ultimately from subducted oceanic crust.

Our study of diamond-hosted minerals from Juina suggests a process whereby subducted, carbonated oceanic crust undergoes low-degree partial melting to produce trace-element-rich carbonatite melts. The mantle transition zone is a plausible location for crustal melting as slabs may founder owing to the density crossover between eclogite and peridotite<sup>17,33</sup>, and could heat up to temperatures of the carbonated eclogite solidus. The distinctive chemical 'flavour' of primary carbonatites, particularly their oceanic crustal signatures and extreme enrichment in trace elements, ensures that the consequences of their passage are preserved for very long timescales, even in rocks that have long since lost any carbonate component through low-pressure decarbonation reactions. As metasomatic agents, carbonatites in effect impart a 'stain' on mantle rocks that can persist for billions of years, and which can in turn impart a unique chemical and isotopic signature to mantle-derived magmas erupted at the Earth's surface.



**Figure 4 | Relative compatibility diagrams showing trace element abundances of model primitive carbonatites normalized to primitive mantle.** The light shaded field shows calculated melts that could coexist with calcic majorite garnets from this study and one previously described<sup>31</sup>. The dark shaded field shows estimates of primitive carbonatite compositions from the literature<sup>27,30</sup>, and the intermediate shaded field shows the composition of glasses trapped in mantle xenoliths from ocean island basalts (OIB)<sup>26,28,29</sup>.



## METHODS SUMMARY

We exposed mineral inclusions *in situ* by polishing diamonds along dodecahedral planes. Cathodoluminescence imaging of polished diamond slabs confirmed that the inclusions are not in proximity to cracks, suggestive of syngensis with the diamond hosts. Major and minor element compositions of the inclusions were determined with a Cameca SX100 electron microprobe at the University of Bristol, operated at 20 nA and 15 kV and calibrated against silicate and oxide standards. Trace element concentrations were measured with a Cameca IMS-4f ion-microprobe at the University of Edinburgh using an 11 keV primary beam of  $^{16}\text{O}^-$  ions, a sample current of 2 nA and secondary ion accelerating voltage of 4,500 V, with glass standards as calibrants.

Subsolidus experiments in the system  $\text{CaSiO}_3\text{--CaTiO}_3\text{--MgSiO}_3$  were performed at Bristol in a laser-heated diamond anvil cell. Ground glass starting materials with Pt-black were contained in stainless-steel gaskets with pressure measured by ruby fluorescence. Samples were compressed to 20–50 GPa and heated simultaneously on both sides for 30–60 min at 1,227–2,227 °C; temperature was measured using standard spectroradiometric techniques. Angle dispersive powder diffraction patterns of P–T quenched samples were collected using focused synchrotron X-ray diffraction techniques at the Synchrotron Radiation Source in Daresbury, UK, and the Advanced Light Source in Berkeley, USA. Phases were indexed on the basis of space groups *Pbnm* for Mg-perovskite and *Pm3m* for  $\text{Ca}(\text{Ti,Si})\text{O}_3$  perovskite.

A melting experiment in the system  $\text{CaO--MgO--Al}_2\text{O}_3\text{--SiO}_2\text{--TiO}_2\text{--CO}_2$  was made using multi-anvil techniques at Bayerisches Geoinstitut, Germany. Powdered glass and magnesite starting mixture was placed within a platinum capsule and run at 20 GPa and 1,475 °C in an octahedral pressure cell, with 10-mm edge lengths and tungsten carbide cubes with 5-mm corner truncations, using a 1,200 ton press. Electron microprobe analyses of run products were made with a JEOL JXA-8900 microprobe at Bayerisches Geoinstitut using silicate and oxide standards, a beam current of 15 nA and a 15 kV voltage.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 13 January; accepted 28 May 2008.**

- Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
- Zindler, A. & Hart, S. R. Chemical geodynamics. *Annu. Rev. Earth Planet. Sci.* **14**, 493–571 (1986).
- Haggerty, S. E. in *Mantle Petrology: Field Observations and High Pressure Experimentation* (eds Fei, Y., Bertka, C. M. & Mysen, B. O.) 105–123 (Geochemical Society Special Publications, 1999).
- Harte, B. & Harris, J. W. Lower mantle mineral association preserved in diamonds. *Miner. Mag.* **58**, 384–385 (1994).
- Harte, B., Harris, J. W., Hutchison, M. T., Watt, G. R. & Wilding, M. C. in *Mantle Petrology: Field observations and High Pressure Experimentation* (eds Fei, Y., Bertka, C. M. & Mysen, B. O.) 125–153 (Geochemical Society Special Publications, Houston, 1999).
- Kaminsky, F. *et al.* Superdeep diamonds from the Juina area, Mato Grosso State, Brazil. *Contrib. Mineral. Petrol.* **140**, 734–753 (2001).
- Stachel, T. Diamonds from the asthenosphere and the transition zone. *Eur. J. Mineral.* **13**, 883–892 (2001).
- Tappert, R. *et al.* Subducting oceanic crust: The source of deep diamonds. *Geology* **33**, 565–568 (2005).
- Brenker, F. E. *et al.* Detection of a Ca-rich lithology in the Earth's deep (>300 km) convecting mantle. *Earth Planet. Sci. Lett.* **236**, 579–587 (2005).
- Hayman, P. C., Kopylova, M. G. & Kaminsky, F. V. Lower mantle diamonds from Rio Soriso (Juina area, Mato Grosso, Brazil). *Contrib. Mineral. Petrol.* **149**, 430–445 (2005).
- Bulanova, G. P. The formation of diamond. *J. Geochem. Exp.* **53**, 1–23 (1995).
- Brenker, F. E. *et al.* Carbonates from the lower part of transition zone or even the lower mantle. *Earth Planet. Sci. Lett.* **260**, 1–9 (2007).
- Kubo, A., Suzuki, T. & Akaogi, M. High pressure phase equilibria in the system  $\text{CaTiO}_3\text{--CaSiO}_3$ : stability of perovskite solid solutions. *Phys. Chem. Mineral.* **24**, 488–494 (1997).
- Stachel, T., Harris, J. W., Brey, G. & Joswig, W. Kankan diamonds (Guinea) II: lower mantle inclusion paragenesis. *Contrib. Mineral. Petrol.* **140**, 16–27 (2000).
- Hirose, K. & Fei, Y. Subsolidus and melting phase relations of basaltic composition in the uppermost lower mantle. *Geochim. Cosmochim. Acta* **66**, 2099–2108 (2002).
- Hirose, K., Shimizu, N., vanWestrenan, W. & Fei, Y. Trace element partitioning in Earth's lower mantle and implications for geochemical consequences of partial melting at the core–mantle boundary. *Phys. Earth Planet. Inter.* **146**, 249–260 (2004).
- Irifune, T. & Ringwood, A. E. Phase transformations in subducted oceanic crust and buoyancy relationships at depths of 600–800 km in the mantle. *Earth Planet. Sci. Lett.* **117**, 101–110 (1993).
- McDonough, W. F. & Sun, S.-s. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Safonov, O. G., Perchuk, L. L. & Litvin, Y. A. Melting relations in the chloride–carbonate–silicate systems at high-pressure and the model for formation of alkalic diamond-forming liquids in the upper mantle. *Earth Planet. Sci. Lett.* **253**, 112–128 (2007).
- Arima, M., Kozai, Y. & Akaishi, M. Diamond nucleation and growth by reduction of carbonate melts under high-pressure and high-temperature conditions. *Geology* **30**, 691–694 (2002).
- Keshav, S., Gudfinnsson, G. & Presnall, D. Majoritic-garnets and clinopyroxenes in cratonic diamonds: Precipitates from  $\text{CO}_2$ -rich melts. *Proc. 11th Int. Conf. EMPG abstr.* **36** (2006).
- Corgne, A. & Wood, B. J.  $\text{CaSiO}_3$  and  $\text{CaTiO}_3$  perovskite–melt partitioning of trace elements: implications for gross mantle differentiation. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL014398 (2002).
- Brenan, J. M., Shaw, H. F., Ryerson, F. J. & Phinney, D. L. Mineral-aqueous fluid partitioning of trace elements at 900 °C and 2.0 GPa: Constraints on the trace element chemistry of mantle and deep crustal fluids. *Geochim. Cosmochim. Acta* **59**, 3331–3350 (1995).
- Kessel, R., Schmidt, M., Ulmer, P. & Pettke, T. Trace element signature of subduction-zone fluids, melts and supercritical liquids at 120–180 km depth. *Nature* **437**, 724–727 (2005).
- Manning, C. E. The chemistry of subduction-zone fluids. *Earth Planet. Sci. Lett.* **223**, 1–16 (2004).
- Coltorti, M., Bonadiman, C., Hinton, R. W., Siena, F. & Upton, B. G. J. Carbonatite metasomatism of the oceanic upper mantle: Evidence from clinopyroxenes and glasses in ultramafic xenoliths of Grande Comore, Indian Ocean. *J. Petrol.* **40**, 133–165 (1999).
- Harmer, R. E. & Gittins, J. The case for primary, mantle-derived carbonatite magma. *J. Petrol.* **39**, 1895–1903 (1998).
- Hauri, E., Shimizu, N., Dieu, J. J. & Hart, S. R. Evidence for hotspot-related carbonatite metasomatism in the oceanic upper mantle. *Nature* **365**, 221–227 (1993).
- Ionov, D. A. Trace element composition of mantle-derived carbonates and coexisting phases in peridotite xenoliths from alkali basalts. *J. Petrol.* **39**, 1931–1941 (1998).
- Kogarko, L. N. Geochemical characteristics of oceanic carbonatites from Cape Verde Islands. *S. Afr. J. Geol.* **96**, 119–125 (1993).
- Stachel, T., Brey, G. & Harris, J. W. Kankan diamonds (Guinea) I: from the lithosphere down to the transition zone. *Contrib. Mineral. Petrol.* **140**, 1–15 (2000).
- Dasgupta, R., Hirschmann, M. M. & Withers, A. C. Deep global cycling of carbon constrained by the solubility of anhydrous, carbonated eclogite under upper mantle conditions. *Earth Planet. Sci. Lett.* **227**, 73–85 (2004).
- Hirose, K., Fei, Y., Ma, Y. & Mao, H.-K. The fate of subducted basaltic crust in the Earth's lower mantle. *Nature* **397**, 53–56 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Diamond samples from Collier 4 were collected by Rio Tinto (Rio Tinto Desenvolvimentos Minerais Ltda) in 1994. We thank Rio Tinto for access to the collection and J. Pickles for technical assistance. This work was supported by an NERC grant to M.J.W. Experiments by L.S.A. at Bayerisches Geoinstitut were supported by the Marie Curie 6th Framework Programme. Synchrotron experiments at Synchrotron Radiation Source, Daresbury Laboratory, UK, and at the Advanced Light Source, Berkeley, USA, were supported by awards to M.J.W. Trace element analyses at the NERC Edinburgh Ion Microprobe Facility were supported by an award to M.J.W.

**Author Contributions** M.J.W., G.P.B., J.D.B. and C.B.S. formulated the project. M.J.W., L.S.A., S.K., G.G., O.T.L., A.R.L. and S.M.C. were responsible for experimental and analytical data collection. G.P.B. was responsible for diamond sample preparation. L.G. processed the kimberlite to recover diamonds and selected inclusion-bearing stones for the project. M.J.W. wrote the manuscript with assistance from G.P.B., L.S.A., S.K., J.D.B., A.R.L. and C.B.S.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.J.W. ([m.j.walter@bris.ac.uk](mailto:m.j.walter@bris.ac.uk)).

## METHODS

The diamonds are 2–4 mm colourless octahedron/dodecahedron transition forms. Diamonds J9 and J10 are whole crystals; stone J1 is mostly whole but with minor breakage along the dodecahedral plane. The diamond faces are resorbed, heavily etched and the broken surface of the J1 crystal is also covered by natural etch patterns. Observation of the whole stones in photoluminescence light shows their dark-blue colour and lack of prominent zonation. Pyrrhotite inclusions are present in J1 and J10 stones in addition to silicates and are surrounded by a rosette of local black cracks. Optical microscopy of the unpolished diamond did not show any cracks associated with the silicate inclusions. Cathodoluminescence imaging of polished diamond plates with the CaTiSi-phase exposed shows that the inclusions have no cracks going to the surface and are located deep within the diamonds, consistent with syngenetic growth with the diamond hosts. Further, inclusion morphologies show imposed negative diamond shapes, also supporting a syngenetic interpretation. The same applies to the majorite in J9 stone. On the basis of the cathodoluminescence imaging, the J1 garnet inclusion is located either in a small diamond intergrowth or in a small diamond block surrounded by an ancient re-healed crack.

Electron microprobe analysis of diamond inclusions was made at University of Bristol with a Cameca SX100 using silicate and oxide standards. We used a beam current of 20 nA and a 15 kV voltage. Conventional ZAF data reduction techniques were used. Replicate analysis of standards yielded uncertainties at the 2% and 5% level for major and minor elements, respectively. Trace elements were determined using a Cameca IMS-4f ion-microprobe at the NERC Edinburgh Ion Microprobe Facility, University of Edinburgh. The primary beam was  $\sim 11$  keV  $^{16}\text{O}^-$  ions ( $\sim 15$  keV net impact energy), with a sample current of 2 nA corresponding to a spatial resolution of  $\sim 15$   $\mu\text{m}$  at the sample surface. The secondary ion accelerating voltage of 4,500 V was offset by 75 eV (energy window of 40 eV) to reduce molecular ion transmission. Calibration was performed on glass standards under identical operating conditions. Statistical precision is better than 10% relative for all isotopes. Accuracy is better than 10% relative for REE barium, strontium, niobium, zirconium and yttrium. Hafnium, rubidium, thorium and U are accurate to within 30% relative.

Subsolidus phase relations were determined in the system  $\text{CaSiO}_3$ – $\text{CaTiO}_3$ – $\text{MgSiO}_3$  at pressures of 20 to 50 GPa and 1,500 to 2,500 K using laser-heated diamond anvil cell techniques at University of Bristol<sup>34</sup>. Powdered-glass starting mixtures (doubly-fused and triply ground) with 10 wt% Pt-black as a laser

absorber were loaded into  $\sim 100$   $\mu\text{m}$  holes drilled in pre-indented stainless steel gaskets. Samples were compressed in diamond anvil cells using diamonds with 250 or 300  $\mu\text{m}$  culets. Pressure was measured by ruby fluorescence. Samples were heated for 30 to 60 min using a double-sided heating geometry, with temperature measurement using standard radiometric techniques<sup>35</sup>. Phase identification in P–T quenched samples was determined using focused synchrotron X-ray diffraction at station 9.5 HPT of the SRS, Daresbury Laboratory, UK, and station 12.2.2 at the ALS, Lawrence Berkeley Laboratory, USA. At the SRS a final pair of slits with a rectangular geometry of  $\sim 20 \times 40$   $\mu\text{m}$  truncate the focused X-ray beam. At the ALS, Kirkpatrick-Baez mirrors are used to focus the X-ray beam to a  $\sim 20$   $\mu\text{m}$  diameter spot. We routinely interrogate the central part of the  $\sim 80$   $\mu\text{m}$  radius heated region in several acquisitions (300–900 s per acquisition). Wavelength dispersive diffraction spectra are acquired with MAR345 imaging plates at both facilities and data are reduced to intensity– $2\theta$  plots using FIT2D software. Mg-perovskite and  $\text{Ca}(\text{Ti,Si})\text{O}_3$ -perovskites are indexed according to *Pbnm* and *Pm3m* crystal structures. Full details of all experimental compositions, run products and phase relation are to be presented elsewhere.

A melting experiment in the system  $\text{CaO}$ – $\text{MgO}$ – $\text{Al}_2\text{O}_3$ – $\text{SiO}_2$ – $\text{TiO}_2$ – $\text{CO}_2$  in which Ca-rich carbonate melt coexists with majorite, Ca-perovskite and magnesite was made at 20 GPa and 1,475 °C using multi-anvil techniques at Bayerisches Geoinstitut (BGI), Germany. A model eclogite starting mixture made of powdered glass and magnesite ( $\text{MgCO}_3$ ) was placed within a platinum capsule and run at high pressure and temperature in a 10/5-type pressure cell using a 1,200 ton press and a ‘Kawai’-type anvil geometry. The sample was held at temperature for 30 min and then quenched. Electron microprobe analyses of the mineral and melt phase in the run product was made at BGI with a JEOL JXA-8900 microprobe using silicate and oxide standards, a beam current of 15 nA and a 15 kV voltage with conventional ZAF data reduction techniques. Full details are to be presented elsewhere.

34. Walter, M. J. *et al.* Subsolidus phase relations and perovskite compressibility in the system  $\text{MgO}$ – $\text{AlO}_{1.5}$ – $\text{SiO}_2$  with implications for Earth's lower mantle. *Earth Planet. Sci. Lett.* **248**, 77–89 (2006).
35. Walter, M. J. & Koga, K. T. The effects of chromatic dispersion on temperature measurement in the laser-heated diamond anvil cell. *Phys. Earth Planet. Inter.* **143–144**, 541–558 (2004).

## LETTERS

# Environmental determinants of extinction selectivity in the fossil record

Shanan E. Peters<sup>1</sup>

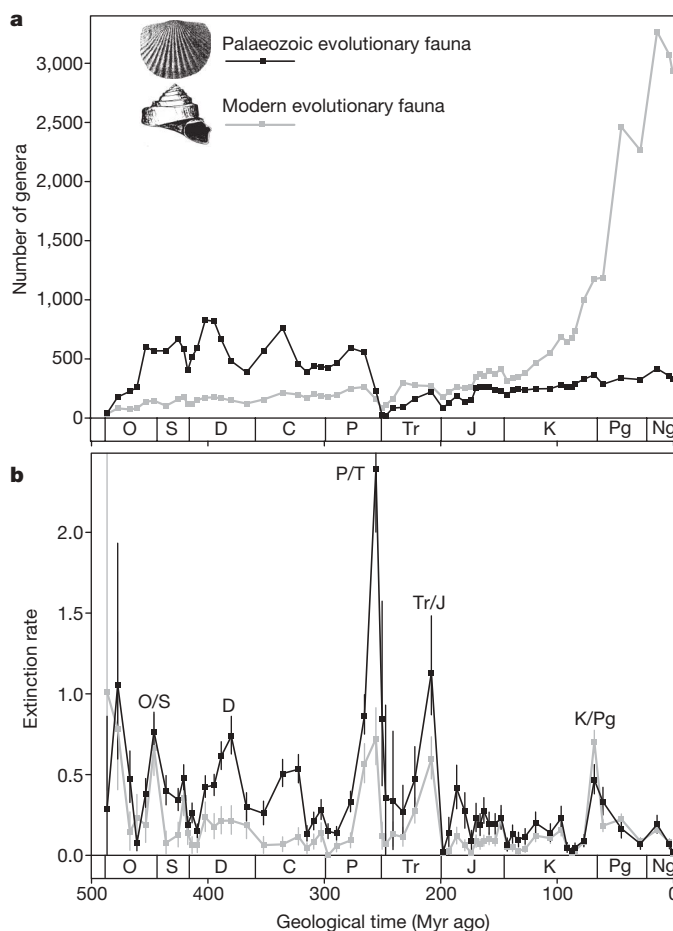
The causes of mass extinctions and the nature of biological selectivity during extinction events remain central questions in palaeobiology. Although many different environmental perturbations have been invoked as extinction mechanisms<sup>1–3</sup>, it has long been recognized that fluctuations in sea level coincide with many episodes of biotic turnover<sup>4–6</sup>. Recent work supports the hypothesis that changes in the areas of epicontinental seas have influenced the macroevolution of marine animals<sup>7,8</sup>, but the extent to which differential environmental turnover has contributed to extinction selectivity remains unknown. Here I use a new compilation of the temporal durations of sedimentary rock packages to show that carbonate and terrigenous clastic marine shelf environments have different spatio-temporal dynamics and that these dynamics predict patterns of genus-level extinction, extinction selectivity and diversity among Sepkoski's Palaeozoic and modern evolutionary faunae<sup>9</sup>. These results do not preclude a role for biological interactions or unusual physical events as drivers of macroevolution, but they do suggest that the turnover of marine shelf habitats and correlated environmental changes have been consistent determinants of extinction, extinction selectivity and the shifting composition of the marine biota during the Phanerozoic eon.

One of the most striking features of the Phanerozoic fossil record of marine animals is a shift in the relative taxonomic richness of clades that first appeared early in the Palaeozoic era. Sepkoski summarized this transition by identifying three evolutionary faunae, each composed of Linnaean classes with covarying diversity trajectories<sup>9</sup>, broadly similar ecologies<sup>9–11</sup>, and comparable mean rates of taxonomic turnover<sup>12</sup>. The Palaeozoic evolutionary fauna, which is typified by low-biomass, epibenthic suspension feeders, was dominant during the Palaeozoic (Fig. 1a) and has higher and more variable rates of extinction (Fig. 1b) than the modern evolutionary fauna, which is typified by mobile infaunal suspension feeders and grazers<sup>9–12</sup> (see Methods). Selective extinction with respect to evolutionary fauna is prominent at the end-Permian mass extinction, but other time intervals also show differential rates of extinction (Fig. 1b).

Explanations for the transition between the Palaeozoic and modern evolutionary faunae have focused on biological interactions. For example, Sepkoski modelled evolutionary fauna diversity dynamics using coupled logistic equations that assume negative ecological interactions<sup>12</sup>, but tests of this hypothesis fail at the genus level<sup>13,14</sup>. The diversification of the modern evolutionary fauna and the decline of the Palaeozoic evolutionary fauna has also been linked to increasing predation intensity during the Phanerozoic<sup>15,16</sup>, but this hypothesis is not supported by genus occurrence data<sup>17</sup>, nor does it predict the observed extinction selectivity (Fig. 1b). High rates of Palaeozoic evolutionary fauna extinction at the end-Permian mass extinction have been attributed to ecological factors and causally linked to the rise of the modern evolutionary fauna<sup>12,13,18–20</sup>, but this does not

explain longer-term shifts in evolutionary fauna dominance or selective extinction in other time intervals (Fig. 1).

Here I explore the hypothesis that patterns of extinction and extinction selectivity among the evolutionary faunae are linked, by means of differential environmental tolerances, to spatio-temporal dynamics in two types of marine shelf environments: those with sediments derived from the physical erosion of land (siliciclastics), and those with sediments precipitated as calcium carbonate (carbonates),



**Figure 1 | Palaeozoic and modern evolutionary fauna genus diversity and extinction.** **a**, Total number of genera. **b**, Per-capita, per-interval extinction rates. Error bars show 95% binomial confidence limits. Major mass extinctions are labelled (O/S, end-Ordovician; D, late-Devonian; P/T, end-Permian; Tr/J, end-Triassic; K/Pg, end-Cretaceous). O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Palaeogene; Ng, Neogene. Myr ago, million years ago.

<sup>1</sup>Department of Geology & Geophysics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.



either inorganically or in association with biological activity, mainly photosynthetic algae<sup>21</sup>. Carbonate and siliciclastic environments are easily recognized in the rock record and differ markedly in terms of average substrate characteristics and nutrient levels<sup>21</sup>, factors that are expected to influence the relative success of the Palaeozoic and modern evolutionary faunae<sup>10,11,22</sup>. The carbonate–siliciclastic distinction is also important because biological preferences for these environments have been linked empirically to different macroevolutionary trajectories<sup>23,24</sup>.

To test the environmental determinant hypothesis, I first used the Paleobiology Database to determine whether or not Sepkoski's evolutionary faunae occur preferentially in carbonates or siliciclastics (Methods; see also Supplementary Information). In agreement with the average ecological characteristics of their constituent genera<sup>9,11</sup>, most Palaeozoic evolutionary fauna genus occurrences (59% of 96,403) are from carbonate environments, which tend to have low nutrients and firm substrates<sup>21</sup>, whereas most modern evolutionary fauna genus occurrences (61% of 134,833) are from siliciclastics, which tend to have higher nutrients and softer substrates. Thus, it is possible that differential turnover between these two environments has contributed to extinction selectivity during the Phanerozoic eon.

To quantify the sedimentary record in a way that permits direct comparison to macroevolutionary patterns, I compiled the temporal ranges of 3,940 gap-bound carbonate and siliciclastic rock packages at 541 locations in the United States (see Methods). The total number of packages is comparable to taxonomic diversity but is an area- and temporal-continuity-weighted measure of rock quantity<sup>25</sup>. Rates of package truncation are analogous to rates of taxonomic extinction<sup>26</sup> but reflect instead area-weighted rates of environmental truncation.

Shelf carbonate packages decline in number from a maximum in the Ordovician to minima in the Triassic and Recent (Fig. 2a). This long-term trajectory was influenced by the northwards drift of Laurentia from tropical to mid- and upper latitudes during the post-Palaeozoic era<sup>27</sup>. However, even after accounting for an overall reduction in tropical shelf area and unequal latitudinal sampling, there remains a significant decline in the global shelf area of carbonates relative to siliciclastics<sup>28</sup>.

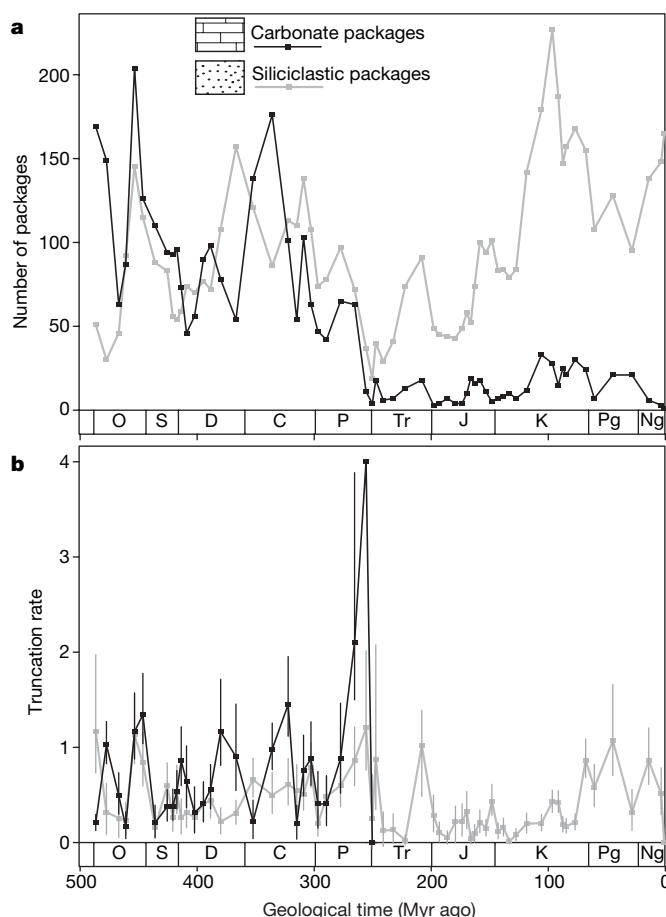
In addition to having different long-term abundance trajectories, carbonates have higher and more variable rates of truncation than siliciclastics (Fig. 2b). This is an important distinguishing characteristic that can be attributed to two main factors: (1) shelf carbonates form over a narrower bathymetric range than siliciclastics; and (2) carbonate environments are subject to negative interactions with siliciclastics because terrigenous sediment input can preclude or dilute carbonate production<sup>21</sup>. Thus, increases in terrigenous sediment supply, which can be driven by sea-level fall, tectonic uplift and climate change, can result in the selective elimination of carbonate environments relative to siliciclastics. Volatility in carbonates may also be enhanced by more direct linkage to climate via temperature and seawater chemistry<sup>18,21</sup>.

A visual comparison of evolutionary fauna macroevolutionary patterns (Fig. 1) and the macrostratigraphy of their preferred shelf environments (Fig. 2) suggests numerous similarities, and these are borne out statistically. Palaeozoic evolutionary fauna extinction rates, de-trended to emphasize short-term variability, are significantly positively correlated with carbonates but not siliciclastics (Table 1). The opposite is true for the modern evolutionary fauna, although the correlations are generally weaker and both carbonates and siliciclastics may be important during the Palaeozoic. On longer timescales, the replacement of the Palaeozoic evolutionary fauna by the modern evolutionary fauna (Fig. 1a) is mirrored by the temporal replacement of carbonates by siliciclastics (Fig. 2a).

Despite the fact that sedimentary data from North America are being compared to genus data that are ostensibly global, more than 55% of the variance in interval-to-interval changes in Palaeozoic evolutionary fauna extinction rates during the Palaeozoic is predicted by interval-to-interval changes in rates of carbonate truncation

(Fig. 3). Results are similar when only those genera occurring in North America are used to calculate evolutionary fauna rates (Table 1), when genus extinction is adjusted to account for variable and incomplete sampling<sup>26</sup> (Table 1), when variation in interval duration is factored out via partial correlation, when correlations are calculated for arbitrary stretches of geological time, and when analysis is confined to the Palaeozoic (see Supplementary Information). Thus, not only is the prominent biological selectivity at the end-Permian mass extinction reproduced by the selective truncation of carbonate environments, but it is reproduced in a way that is statistically robust and consistent with the environmental and biological selectivity observed in many other time intervals, including the end-Ordovician, late Devonian and mid-Carboniferous extinctions.

The ability of the sedimentary rock record to predict macroevolutionary patterns might seem to constitute *prima facie* evidence for a geologically controlled sampling bias in fossil data, but three observations allow this hypothesis to be discounted. First, adjusting genus extinction rates to account for variable and incomplete preservation<sup>26</sup> does not change the overall environment–evolutionary fauna relationship (Table 1). Second, analyses of the aggregate sedimentary and fossil records show that sedimentary hiatus durations do not predict rates of genus extinction<sup>8</sup>, and hiatus durations are similarly uncorrelated with evolutionary fauna extinction rates. Third, if sampling bias was a dominant factor, correlations between sedimentary and fossil data would be comparable for all groups of taxa, yet extinction in the Palaeozoic evolutionary fauna is more strongly correlated with



**Figure 2 | Carbonate and siliciclastic macrostratigraphy.** **a**, Total number of packages, an area- and temporal-continuity-weighted measure of rock quantity. **b**, Per-package, per-interval truncation rates, an area-weighted measure of environmental truncation. Error bars show 95% binomial confidence limits. There are too few carbonate packages in the post-Palaeozoic to reliably calculate truncation rates.

**Table 1 | Spearman rank-order correlation coefficients**

| Lithofacies                   | Raw extinction rates | Corrected extinction rates | North America extinction rates |
|-------------------------------|----------------------|----------------------------|--------------------------------|
| Palaeozoic evolutionary fauna |                      |                            |                                |
| Carbonate*                    | 0.752‡               | 0.686‡                     | 0.747‡                         |
| Siliciclastic                 | 0.204 (NS)           | 0.106 (NS)                 | 0.157 (NS)                     |
| Modern evolutionary fauna     |                      |                            |                                |
| Carbonate*                    | 0.265 (NS)           | 0.061 (NS)                 | 0.547†                         |
| Siliciclastic                 | 0.409†               | 0.451†                     | 0.406†                         |

Data are de-trended (first differences) time series of genus extinction rates and environmental truncation rates. Raw extinction rates are face-value extinction rates in Sepkoski's compendium (Fig. 1b). Corrected extinction rates are extinction rates corrected for variable and incomplete sampling (mean of 194 pulsed model iterations<sup>26</sup>). North America extinction rates use only genera occurring in North America according to the Paleobiology Database (~30% of genera used to calculate raw extinction rates). Cross-correlations for carbonate-siliciclastic, 0.116\* (NS); Palaeozoic-modern evolutionary faunae, 0.675 ( $P < 0.0001$ ). NS, not significant ( $P \geq 0.1137$ ).

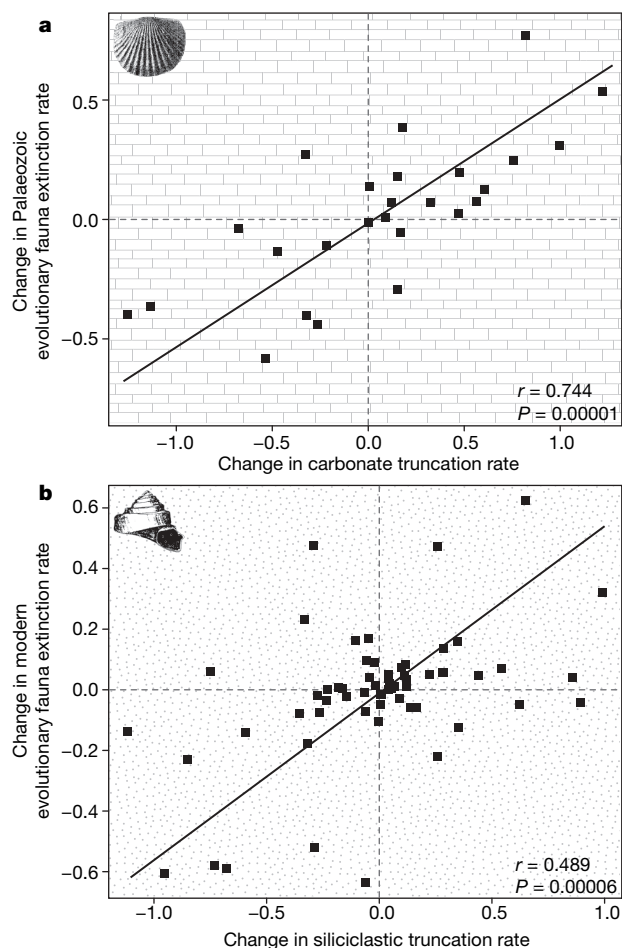
\* Carbonate rates are constrained only in the Palaeozoic (Tremadocian-Guadalupian;  $n = 27$ ). The last Permian stage (Tatarian) is omitted because of limited carbonate packages; including it strengthens the correlation between carbonate and Palaeozoic evolutionary fauna (raw extinction rates  $r = 0.778$ ). Correlations for siliciclastics include Tremadocian-Pliocene ( $n = 62$ ).

†  $P \leq 0.0043$ .

‡  $P \leq 0.0002$ .

carbonate truncation than the modern evolutionary fauna is correlated with either environment.

The inability of sampling bias to explain adequately environment-evolutionary fauna correlations leaves two alternative hypotheses.



**Figure 3 | First differences in evolutionary fauna extinction rates versus environmental truncation rates.** **a**, Palaeozoic evolutionary fauna versus carbonate. **b**, Modern evolutionary fauna versus siliciclastic. Linear product-moment correlation coefficients,  $P$ -values and reduced major axis trend lines are shown. The last Permian stage is omitted in **a** owing to low numbers of carbonate packages (if included,  $r = 0.815$ ). See Table 1 for rank-order correlations for all pairwise comparisons.

First, it is possible that genus last occurrences are taxonomic artefacts, and that the observed correlations reflect a tendency among systematists to arbitrarily truncate continuous lineage segments during times of environmental turnover. Although conceivable, this hypothesis is unlikely because short-term variability in rates of extinction would persist even if most lineages were artificially truncated, provided that true lineage extinctions also result in genus last occurrences (see Supplementary Information). The second, much more probable, hypothesis is that both evolutionary fauna turnover and shelf sedimentation share a common cause. The most plausible mechanism is sea-level change and the resultant expansion and contraction of epicontinental seas, phenomena which are intimately linked to tectonics via crustal uplift and subsidence, and to global climate via temperature and continental ice volume.

Whatever the reasons for the close statistical similarities between patterns of marine shelf sedimentation and rates of extinction among Sepkoski's Palaeozoic and modern evolutionary faunae, these results must affect our understanding of large-scale patterns in the fossil record. If the correlations reflect taxonomic artefacts, then extinction rate estimates, including those at the major mass extinctions, require substantial revision. If, instead, both selective biological extinction and differential environmental turnover share a common cause, then mechanisms for extinction and extinction selectivity gain important new constraints. Although the ultimate explanation is likely to be more complicated than a simple species area effect, these results do provide a substantial measure of support for the intriguing and long-standing hypothesis<sup>5,6</sup> that changes in the areas of unique epicontinental sea habitats, as well as correlated environmental effects, have consistently influenced rates of extinction, extinction selectivity, and the shifting composition of the marine biota during the Phanerozoic eon.

## METHODS SUMMARY

Genus data derive from Sepkoski's global compendium<sup>29</sup> of marine animal genera (<http://strata.geology.wisc.edu/jack>). Only genera with both range end-points resolved to one of Sepkoski's 63 post-Cambrian time intervals (stages) were used. The evolutionary faunae were based on Sepkoski<sup>9</sup> and partitioned as follows. A total of 8,009 genera from Bivalvia, Echinoidea, Gastropoda, Gymnolaemata and Malacostraca constitute the modern evolutionary fauna, and 7,460 genera from Anthozoa, Articulata, Asteroidea, Blastoidea, Crinoidea and Stenolaemata comprise the Palaeozoic evolutionary fauna. Together, these two evolutionary faunae include more than 70% of all post-Cambrian genera in Sepkoski's compilation.

To estimate environmental preferences, global genus occurrences were downloaded from the Paleobiology Database (<http://paleodb.org>) marine invertebrate working group on 7 September 2007. Multiple species from the same genus in a collection were grouped together and subgenera were kept separate. Of 308,865 genus occurrences, 256,481 (83%) could be assigned to one of three evolutionary faunae. Each collection was assigned to one of four lithologies (carbonate, siliciclastic, mixed, unknown) on the basis of the primary and secondary lithology fields (see Supplementary Information). Occurrences in mixed collections (those with primary and secondary lithologies from two different environments) were assigned one-half of an occurrence to each environment. Changing this protocol does not substantively affect the results.

Macrostratigraphy is based on the temporal ranges of gap-bound rock packages compiled separately for different geographical locations<sup>25</sup>. The correlation charts of ref. 30 were used to compile the temporal ranges of surface and subsurface rock packages at 541 locations in the United States (see Supplementary Information). A total of 1,518 carbonate and 2,422 siliciclastic packages were recognized. Reference 26 describes the survivorship-based rate parameters calculated for both gap-bound rock packages and marine animal genera.

Full Methods and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 6 February; accepted 28 April 2008.

Published online 15 June 2008.

- Bambach, R. K. Phanerozoic biodiversity mass extinctions. *Annu. Rev. Earth Planet. Sci.* **34**, 127–155 (2006).
- Hallam, A. & Wignall, P. B. *Mass Extinctions and their Aftermath* (Oxford, Oxford, 1997).

3. Raup, D. M. Large-body impact and extinction in the Phanerozoic. *Paleobiology* **18**, 80–82 (1992).
4. Hallam, A. The case for sea-level change as a dominant causal factor in mass extinction of marine invertebrates. *Proc. R. Soc. Lond. B* **325**, 437–455 (1989).
5. Newell, N. D. Periodicity in invertebrate paleontology. *J. Paleontol.* **26**, 371–385 (1952).
6. Johnson, J. G. Extinction of perched faunas. *Geology* **2**, 479–482 (1974).
7. Peters, S. E. Geologic constraints on the macroevolutionary history of marine animals. *Proc. Natl Acad. Sci. USA* **102**, 12326–12331 (2005).
8. Peters, S. E. Genus extinction, origination, and the durations of sedimentary hiatuses. *Paleobiology* **32**, 387–407 (2006).
9. Sepkoski, J. J. Jr. A factor analytic description of the Phanerozoic marine fossil record. *Paleobiology* **7**, 36–53 (1981).
10. Bambach, R. K. in *Phanerozoic Diversity Patterns: Profiles in Macroevolution* (ed. Valentine, J. W.) 191–253 (Princeton, Princeton, 1985).
11. Bambach, R. K. Seafood through time: changes in biomass, energetics, and productivity in the marine ecosystem. *Paleobiology* **19**, 372–397 (1993).
12. Sepkoski, J. J. Jr. A kinetic-model of Phanerozoic taxonomic diversity 3: post-Paleozoic families and mass extinctions. *Paleobiology* **10**, 246–267 (1984).
13. Stanley, S. M. An analysis of the history of marine animal diversity. *Paleobiology* **33** (Suppl.), 1–55 (2007).
14. Alroy, J. Are Sepkoski's evolutionary faunas dynamically coherent? *Evol. Ecol. Res.* **6**, 1–32 (2004).
15. Stanley, S. M. in *Patterns of Evolution as Illustrated by the Fossil Record* (ed. Hallam, A.) 209–250 (Elsevier, Amsterdam, 1977).
16. Vermeij, G. J. *Evolution and Escalation* (Princeton, Princeton, 1987).
17. Madin, J. S. *et al.* Statistical independence of escalatory ecological trends in Phanerozoic marine invertebrates. *Science* **312**, 897–900 (2006).
18. Knoll, A. H., Bambach, R. K., Payne, J. L., Pruss, S. & Fischer, W. W. Paleophysiology and the end-Permian mass extinction. *Earth Planet Sci. Lett.* **256**, 295–313 (2007).
19. Fraiser, M. L. & Bottjer, D. J. When bivalves took over the world. *Paleobiology* **33**, 397–413 (2007).
20. Erwin, D. E. *Extinction: How Life on Earth Nearly Ended 250 Million Years Ago* (Princeton, Princeton, 2006).
21. Wright, V. P. & Burchette, T. P. in *Sedimentary Environments* (ed. Reading, H. G.) 325–394 (Blackwell, Oxford, 1996).
22. McKinney, F. K. & Hageman, S. J. Paleozoic to modern marine ecological shift displayed in the northern Adriatic Sea. *Geology* **34**, 881–884 (2006).
23. Foote, M. Substrate affinity and diversity dynamics of Paleozoic marine animals. *Paleobiology* **32**, 345–366 (2006).
24. Kiessling, W. & Aberhan, M. Environmental determinants of marine benthic biodiversity dynamics through Triassic–Jurassic time. *Paleobiology* **33**, 414–434 (2007).
25. Peters, S. E. Macrostratigraphy of North America. *J. Geol.* **114**, 391–412 (2006).
26. Foote, M. Origination and extinction components of taxonomic diversity: a new approach. *J. Geol.* **111**, 1125–1148 (2003).
27. Allison, P. A. & Briggs, D. E. G. Paleolatitudinal sampling bias, Phanerozoic species-diversity, and the end-Permian extinction. *Geology* **21**, 65–68 (1993).
28. Walker, L., Wilkinson, B. & Ivany, L. C. Continental drift and Phanerozoic carbonate accumulation in shallow-shelf and deep-marine settings. *J. Geol.* **110**, 75–87 (2002).
29. Sepkoski, J. J. Jr. *A Compendium of Fossil Marine Animal Genera*. Bull. Am. Paleontol. 363 (Paleontological Research Institution, Ithaca, New York, 2002).
30. Childs, O. E. Correlation of stratigraphic units of North America: COSUNA. *Bull. Am. Assoc. Petrol. Geol.* **69**, 173–180 (1985).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** I thank M. Foote for discussion. M. Foote, W. Kiessling and B. Wilkinson read the manuscript. I also acknowledge the donors of the American Chemical Society and US National Science Foundation EAR-0544941 for financial support.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.E.P. ([peters@geology.wisc.edu](mailto:peters@geology.wisc.edu)).



## METHODS

**Genus data.** Sepkoski's original definition<sup>9</sup> of the evolutionary faunae included nektonic and planktonic taxa, such as nautiloids and vertebrates. Here, only primarily benthic invertebrate classes are assigned an evolutionary fauna (see Methods Summary). Including classes with primarily planktonic or nektonic genera does not significantly alter the results (see Supplementary Information). Because the Palaeozoic and modern evolutionary faunae are not sufficiently taxonomically diverse to calculate rates in the Cambrian, all analyses are limited to the post-Cambrian.

**Sedimentary package data.** Macrostratigraphy is based on the temporal ranges of gap-bound rock packages compiled separately for different geographical locations<sup>25</sup>. For this analysis, temporal gaps in sedimentation fall into one of two categories: (1) those that constitute missing time (hiatuses); and (2) those that constitute missing environment. For example, a unit designated as carbonate but bracketed above and below by siliciclastics would constitute a single gap-bound carbonate package. Similarly, a unit designated as carbonate and bracketed by hiatuses would also constitute a single gap-bound carbonate package (see Supplementary Information). This definition of gaps means that the timing of environmental truncation may be controlled by non-deposition and/or erosional truncation of a longer-duration sediment package. Although the relative frequencies of package termination types cannot be measured at this time, it is likely that most of the environmental gaps recognized here reflect non-deposition rather than significant erosional truncation, if for no other reason than because marine sediments and gaps repeatedly accrue only in areas of low elevation and low net rates of erosion. To be recognized as a significant gap, hiatuses must span approximately one to three million years, although the ability to recognize such gaps varies temporally and spatially according to geological knowledge.

To partition sedimentary packages into carbonates and siliciclastics, the dominant lithologies and environmental data in ref. 30 were followed. Most rock units can be assigned to either carbonates or siliciclastics based on the charts, which reflect the dominant, rather than exclusive, lithology of the specified stratigraphic units. In the few cases where the dominant lithology was reported as mixed carbonate and siliciclastic, the unit was assigned to carbonates, although results are not sensitive to this procedure. It is not always possible to determine from the charts whether a unit was deposited in a marine or terrestrial setting, but in cases where terrestrial sediments could be identified they were excluded. The inclusion of an unknown quantity of terrestrial siliciclastic sediment may be responsible for the lower overall siliciclastic correlations (Table 1).

Although many Phanerozoic carbonates contain abundant animal skeletons, it is important to acknowledge that animal fossils are not required for the formation of carbonate sediments. A substantial fraction of the carbonate in modern shelf environments is produced by photosynthetic algae<sup>21</sup>, and carbonates are abundant in similar shelf settings throughout the Proterozoic, long before animals with skeletons evolved. Thus, animal fossils tend to be found in carbonates that would probably have formed in their absence, just as fossils are found in siliciclastic sediments that would have accumulated with or without skeletal input.

Finally, it is worth noting that both Sepkoski's genus data and the macrostratigraphic data have numerous correlation, categorical and transcription errors. However, such errors should weaken, not inflate, the correlations reported here. Furthermore, the possibility that genus last occurrences are arbitrary subdivisions of continuous lineage segments does not demand that interval-to-interval changes in rates of genus extinction diverge statistically from true underlying rates of extinction, provided that true lineage extinction still results in a significant number of recognized genus last occurrences (see Supplementary Information).

## LETTERS

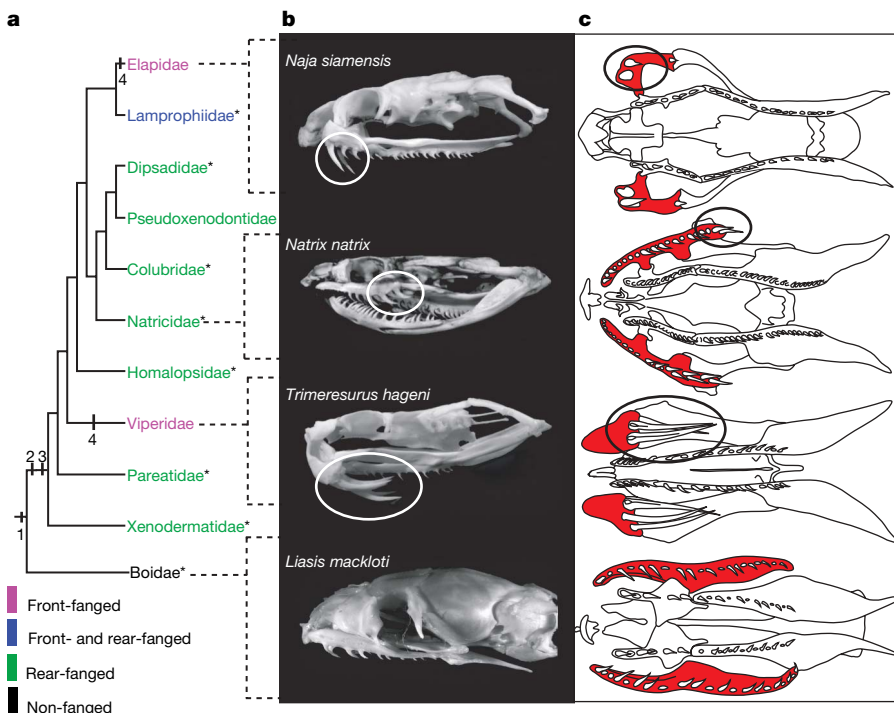
## Evolutionary origin and development of snake fangs

Freek J. Vonk<sup>1</sup>, Jeroen F. Admiraal<sup>1</sup>, Kate Jackson<sup>2</sup>, Ram Reshef<sup>3</sup>, Merijn A. G. de Bakker<sup>1</sup>, Kim Vanderschoot<sup>1</sup>, Iris van den Berge<sup>1</sup>, Marit van Atten<sup>1</sup>, Erik Burgerhout<sup>1</sup>, Andrew Beck<sup>4</sup>, Peter J. Mirtschin<sup>4,5</sup>, Elazar Kochva<sup>6</sup>, Frans Witte<sup>1</sup>, Bryan G. Fry<sup>7</sup>, Anthony E. Woods<sup>4</sup> & Michael K. Richardson<sup>1</sup>

Many advanced snakes use fangs—specialized teeth associated with a venom gland<sup>1,2</sup>—to introduce venom into prey or attacker. Various front- and rear-fanged groups are recognized, according to whether their fangs are positioned anterior (for example cobras and vipers) or posterior (for example grass snakes) in the upper jaw<sup>3–5</sup>. A fundamental controversy in snake evolution is whether or not front and rear fangs share the same evolutionary and developmental origin<sup>3–9</sup>. Resolving this controversy could identify a major evolutionary transition underlying the massive radiation of advanced snakes, and the associated developmental events. Here we examine this issue by visualizing the tooth-forming epithelium in the upper jaw of 96 snake embryos, covering eight species. We use the sonic hedgehog gene as a marker<sup>10–13</sup>, and three-dimensionally reconstruct the development in 41 of the embryos. We show that front fangs develop from the posterior end of the upper jaw, and are strikingly similar in morphogenesis to rear fangs. This is consistent with their being homologous. In front-fanged snakes, the anterior part of the upper jaw lacks sonic hedgehog expression, and ontogenetic allometry displaces the fang

from its posterior developmental origin to its adult front position—consistent with an ancestral posterior position of the front fang. In rear-fanged snakes, the fangs develop from an independent posterior dental lamina and retain their posterior position. In light of our findings, we put forward a new model for the evolution of snake fangs: a posterior subregion of the tooth-forming epithelium became developmentally uncoupled from the remaining dentition, which allowed the posterior teeth to evolve independently and in close association with the venom gland, becoming highly modified in different lineages. This developmental event could have facilitated the massive radiation of advanced snakes in the Cenozoic era, resulting in the spectacular diversity of snakes seen today<sup>6,14,15</sup>.

Many advanced snakes (Caenophidia, in the sense of ref. 16) use venom, with or without constriction, to subdue their prey<sup>6,15</sup>. Their venom-delivery system includes a post-orbital venom gland associated with specialized venom-conducting fangs<sup>2</sup>. Fangs can occupy various positions on the upper jaw, but are always located on the maxilla and never on any other tooth-bearing bone<sup>17</sup> (Fig. 1c).



**Figure 1 | Adult maxillary dentition mapped onto a molecular snake phylogeny to show relative positions of the various fang types.** **a**, Phylogeny from ref. 16. **b**, **c**, Adult skulls (Supplementary Table 4): lateral views (**b**); palate, schematic ventral views (**c**; maxilla coloured, fangs circled). Asterisks indicate species studied by electron microscopy (Supplementary Fig. 5, Supplementary Table 3). The evolutionary changes leading from an unmodified maxillary dentition to the different fang types in advanced snakes are indicated at the nodes: (1) continuous maxillary dental lamina, no specialized subregions—ancestral condition for advanced snakes; (2) evolution of posterior maxillary dental lamina—developmental uncoupling of posterior from anterior teeth; (3) starting differentiation of the posterior teeth with the venom gland; (4) loss of anterior dental lamina and development of front fangs.

<sup>1</sup>Institute of Biology, Leiden University, Kaiserstraat 63, PO Box 9516, 2300 RA, Leiden, The Netherlands. <sup>2</sup>Department of Biology, Whitman College, Walla Walla, Washington 99362, USA. <sup>3</sup>Faculty of Biology, Technion - Israel Institute of Technology, Haifa 32000, Israel. <sup>4</sup>Sansom Institute, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia 5000, Australia. <sup>5</sup>Venom Supplies Pty Ltd, Tanunda, South Australia 5352, Australia. <sup>6</sup>Department of Zoology, Tel Aviv University, Tel Aviv 69978, Israel. <sup>7</sup>Department of Biochemistry & Molecular Biology, Bio21 Institute, University of Melbourne, Parkville, Victoria 3010, Australia.

Viperidae (vipers and pit vipers), *Atractaspis* (Lamprophiidae, in the sense of ref. 16) and Elapidae (cobras and their relatives) have tubular front fangs (Fig. 1b, c). The remaining lineages do not have front fangs, being either 'non-fanged' (no distinguishable enlarged posterior tooth) or rear-fanged<sup>6,17</sup> (Fig. 1b, c). Rear fangs can be solid or slightly or deeply grooved, but are never tubular<sup>17</sup>.

There has been active debate concerning the evolutionary origin of these different fang types, and their relationships to the simple, unmodified teeth of non-fanged<sup>18</sup> basal snakes<sup>3–9,18</sup> such as pythons and boas (Boidae). Proposed hypotheses include the following: (1) front-fanged snakes form a monophyletic group and their fangs are derived from rear fangs<sup>8,19,20</sup>; (2) elapid fangs are derived from front teeth and viperid fangs from rear fangs<sup>21,22</sup>; (3) elapid and viperid fangs are both independently derived from rear fangs<sup>6,7</sup>. Establishing the origin and evolutionary transformation series between these dentition types requires a robust phylogeny to map the characters onto. Because recent molecular phylogenies of advanced snakes place the front-fanged Viperidae as relatively basal and the front-fanged Elapidae as more recently derived<sup>6,16</sup> (Fig. 1a), the current evidence seems to support an 'independent-origin' hypothesis<sup>6</sup>.

So far, this issue has not been examined using a molecular developmental approach, not least because of the difficulty of obtaining snake embryos of all the different species. The development of fangs and venom glands has been studied before in viperids<sup>23,24</sup>, *Natrix*<sup>25,26</sup> (Natricidae, in the sense of ref. 16), *Spalerosophis*, *Thamnophis* and *Telescopus*<sup>1,2</sup> (Colubridae). Those morphological studies identified a common primordium of the venom gland and fangs<sup>1,23,25</sup>, but did not visualize the odontogenic band (tooth-forming epithelium: a band of epithelial tissue that invaginates and forms a dental lamina). Therefore, no conclusions could be drawn about the origin and evolutionary transformation series of the fangs.

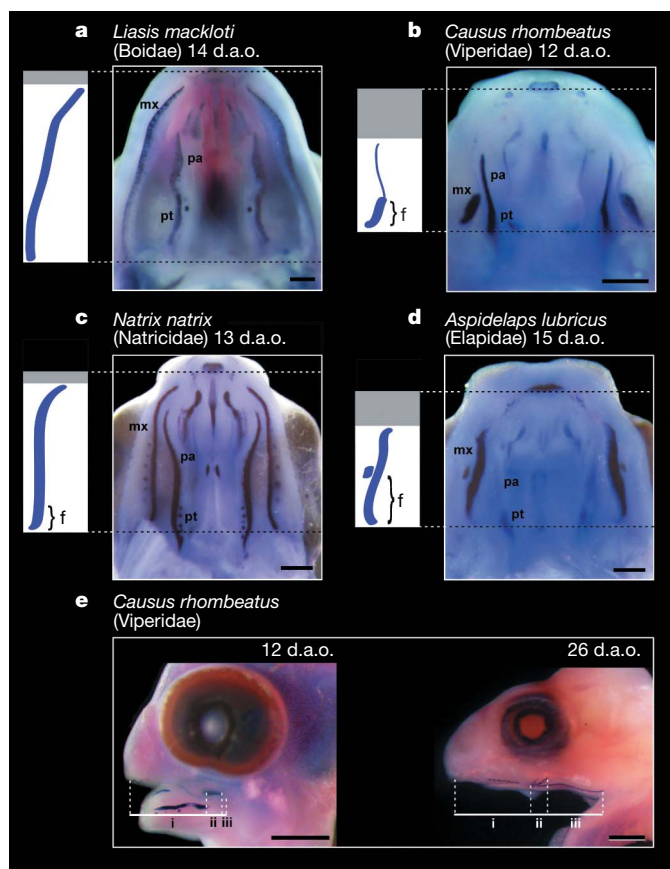
Here we have carried out *in situ* hybridization of the sonic hedgehog (*shh*) gene in 96 snake embryos of multiple stages, and three-dimensionally reconstructed the development of the maxillary dentition in 41 of these through serial sections. We use nine advanced snake species, comprising two front- and two non-front-fanged lineages. As outgroup we included the non-fanged water python, *Liasis mackloti* (Boidae), which is basal to advanced snakes (Fig. 1a). *Shh* is expressed in the odontogenic band in different vertebrate species<sup>10,13</sup>. By visualizing this band, we aim to find evidence for the ancestral condition of the maxillary dentition. A list of all material studied can be found in Supplementary Tables 1–4. We map our characters onto the recently published, robust molecular phylogeny of advanced snakes obtained in ref. 16 (Fig. 1a).

In the water python, *shh* expression reveals one continuous maxillary odontogenic band (Fig. 2a). As confirmed by serial sections of embryos ranging from young to old, this band invaginates to form one dental lamina—a single continuous, invaginating epithelium that will develop a row of teeth (Fig. 3a, Fig. 4j–l). This is consistent with a recent morphological study of *Python sebae*<sup>27</sup>. The odontogenic band, with its associated lamina, appears along the entire rostral–caudal extent of the upper jaw—from the premaxilla to the mandibular articulation (Fig. 2a). This suggests that the ancestral condition for the maxillary dentition of advanced snakes is one dental lamina that appears along the entire rostral–caudal extent of the upper jaw, lacking specialized subregions.

The early odontogenic bands in the non-front-fanged grass snake, *Natrix natrix* (Natricidae), and the rat snake *Elaphe obsoleta* (Colubridae) are similar to that of the water python (Fig. 2c, Supplementary Fig. 2f, i). However, we show that there are two dental laminae which invaginate separately (Figs 3b, e and 4a–c, Supplementary Fig. 3e–g) and fuse during development (Fig. 4c, Supplementary Fig. 3h–i). The anterior lamina bears only teeth (Fig. 4c, Supplementary Fig. 3f) and is similar in development to that in the water python (Fig. 4j–l). The posterior lamina, however, bears teeth and forms the common primordium with a post-orbital gland (Figs 3b and 4b, c; Supplementary Fig. 3i). These develop into the rear

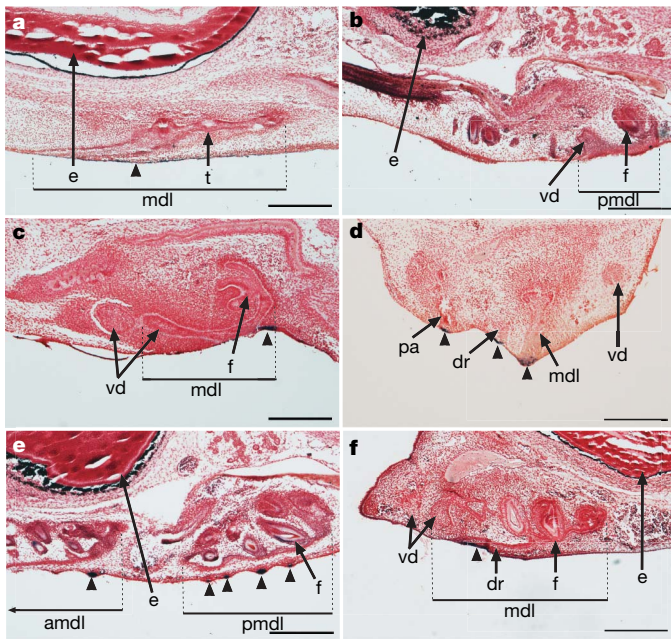
fangs and venom gland in the grass snake, and probably represent the first differentiation of the posterior teeth with a venom gland in the rat snake. The latter observation is consistent with a recent magnetic resonance imaging and histology study<sup>15</sup> showing the presence of a small gland in rat snakes. To verify that the anterior and posterior dental laminae are truly developmentally independent, we ablated the primordium of the anterior lamina in isolated developing upper jaws of the dice snake, *Natrix tessellata*<sup>28</sup> (Supplementary Fig. 3m–o)<sup>29</sup>. We found that, after cultivation under the yolk sac membrane, the posterior lamina, with its venom gland and fangs, developed normally in the absence of the anterior lamina (Supplementary Fig. 4), showing that they are developmentally independent.

In the five front-fanged species examined (Viperidae and Elapidae), the maxillary odontogenic band is found in the posterior part of the upper jaw (Fig. 2b, d; Supplementary Fig. 2a, d, g). There is no *shh* expression or dental lamina in the anterior region (verified by histology; data not shown). In contrast, in the water python, the grass snake and the rat snake, the odontogenic band and associated dental laminae appear along the entire rostral–caudal extent of the upper jaw. We find that, during development, the 'rear' fang is displaced to its adult 'front'



**Figure 2 | *Shh* expression in the embryonic snake palate, showing the posterior developmental origins of front fangs.** **a–d**, Palate, ventral view: top, anterior; scale bar, 0.5 mm; dotted lines, upper jaw (posterior margin of premaxilla to attachment of the mandible); boxes, schemes of maxillary odontogenic band (purple, *shh* expression; grey, no *shh* expression). Positions of fangs in **b–d** were identified histologically (Fig. 3, Supplementary Fig. 3). The odontogenic band in the front-fanged species is located posterior in the upper jaw (**b**, **d**). In the non-fanged outgroup (**a**) and the rear-fanged *Natrix* (**c**), the odontogenic band extends along the entire upper jaw. **f**, fang; **mx**, maxillary odontogenic band; **pa**, palatine odontogenic band; **pt**, pterygoid odontogenic band. **e**, Ontogenetic allometry in the fang in the front-fanged *Causus* displaces the fang along the upper jaw (Supplementary Figs 5–9, Supplementary Tables 5–9). Scale bars, 1 mm. We note the change in relative size of the upper jaw subregions: i, anterior; ii, fang; iii, posterior. d.a.o., days after oviposition.





**Figure 3** | Sections of the *shh* *in situ* hybridizations of the embryonic upper jaw in five snake species, showing the posterior and anterior dental laminae. **a–c, e–f**, Sagittal sections, anterior to the left, of *L. mackloti* (Boidae) 22 d.a.o. (**a**), *N. natrix* (Natricidae) 22 d.a.o. (**b**), *Calloselasma rhodostoma* (Viperidae) 8 d.a.o. (**c**), *N. natrix* 22 d.a.o. (**e**), *Naja siamensis* (Elapidae) 23 d.a.o. (**f**). **d**, Transverse section, medial to the left, of *Trimeresurus hageni* (Viperidae) 8 d.a.o. The posterior maxillary dental laminae in **b** and **e** are similar in morphogenesis to the maxillary dental laminae in all front-fanged species examined (**c, d, f**; see also Fig. 4). Arrowheads, *shh* expression; amdl, anterior maxillary dental lamina; dr, dental ridge; e, eye; f, fang; md, maxillary dental lamina; pa, palatine dental lamina; pmdl, posterior maxillary dental lamina; t, tooth bud; vd, primordium of venom gland; scale bars, 300  $\mu$ m.

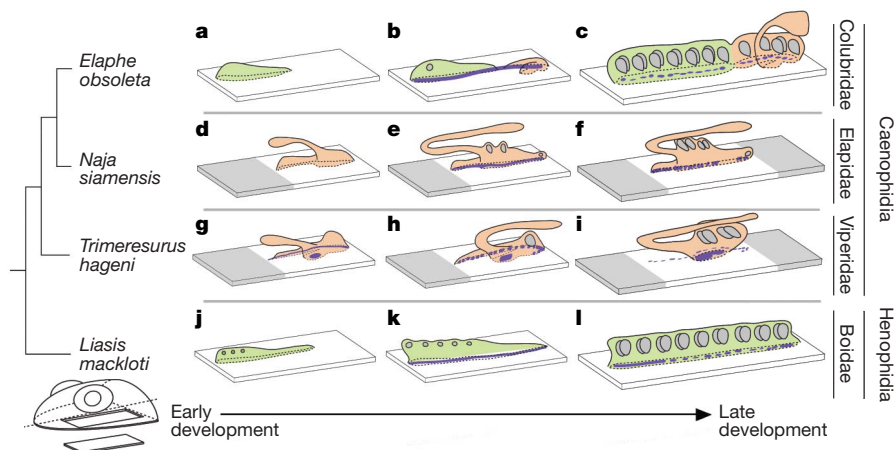
position by ontogenetic allometry (Supplementary Figs 6, 7; Supplementary Table 6 for statistical analyses), suggesting a posterior evolutionary origin for the front fangs. Histology shows that although the odontogenic band invaginates normally and forms one dental lamina

(in contrast to the non-front-fanged snakes described above), in all front-fanged species the fangs develop from the posterior-most part of this lamina and there are no developing teeth in the anterior part (Fig. 3f, Fig. 4d–i). This apparently toothless part of the dental lamina has been described before only in *Vipera aspis* (Viperidae), and termed the ‘dental ridge’<sup>24,30</sup>. We find it here in Elapidae (Fig. 3f, Supplementary Fig. 3j–l). The fact that viperids and elapids share the dental ridge and a posterior developmental origin for their front fangs is interesting, because they are phylogenetically not closely related (Fig. 1a).

Because Elapidae and Viperidae do not form a monophyletic group (Fig. 1a), the dental ridge, the posterior developmental origin of the fangs and the ontogenetic allometry in both lineages may reflect convergent evolution. However, our three-dimensional reconstructions show that there is a striking similarity in morphogenesis of all front and rear fangs examined (Fig. 4b–i), despite the large variation in adult morphology. The toothless dental ridge seen in elapids and viperids is similar to the part of the posterior dental lamina that fuses with the anterior dental lamina in the grass snake and the rat snake (Fig. 4). Although developmental similarity is not conclusive proof of structural homology, this is especially interesting in light of the posterior developmental origin of the front fangs in both elapids and viperids mentioned above. These results are difficult to reconcile with the independent-origin hypothesis, but are consistent with the hypothesis that elapid and viperid front fangs, and the posterior dental lamina in non-front-fanged snakes, represent homologous structures.

Our results suggest a new model for the evolution of snake fangs. A posterior subregion of the ancestral tooth-forming epithelium became developmentally uncoupled from the remaining dentition, resulting in posterior and anterior dental laminae that are developmentally independent (Supplementary Fig. 1). This condition is retained in the non-front-fanged snakes, such as the grass and rat snake. This model would imply that the front-fanged elapids and viperids have independently lost the anterior dental lamina (Fig. 1), which is supported by the lack of *shh* expression anterior in their upper jaws.

Because obtaining developmental data for each non-front-fanged advanced snake lineage is impracticable, convergence cannot be ruled out completely. We have, therefore, examined the adult maxillary tooth morphologies through scanning electron microscopy in the water python, the grass snake, the rat snake and a wide range of other non-front-fanged advanced snake species (Fig. 1a). We aimed to



**Figure 4** | Schematic three-dimensional reconstructions showing the similarity in morphogenesis between the rear and front fangs. Derived from serial sections (Fig. 3, Supplementary Fig. 3); materials analysed are listed in Supplementary Tables 1, 2, 4. Left-hand side of the upper jaw is depicted, and only epithelial components are shown. Purple, *shh* expression; grey, tooth buds; green, unspecialized maxillary dental lamina; orange, specialized maxillary dental lamina that bears fangs. The specialized dental lamina is dilated into a bifurcated epithelial sac, the lateral part giving rise to the venom duct and venom gland by growing rostrad (see also Fig. 3b–d, f),

then turning caudad to reach the post-orbital region (as previously described for vipers<sup>25,26</sup>, *Natrix*<sup>25,26</sup> and *Spalerosophis*<sup>1,2</sup>). In *Elaphe obsoleta* (**a–c**) and *Natrix natrix* (data not shown), fangs develop rostrally and caudally alongside the base of the venom duct; in *Naja siamensis* (**d–f**) and *Trimeresurus hageni* (**g–i**) the rostral part regresses, remaining visible only as the dental ridge, whereas in **b** and **c** this part bears fangs and fuses with the anterior dental lamina. The unspecialized dental lamina in *E. obsoleta* (**a–c**) and the outgroup *Liasis mackloti* (**j–l**) starts developing anterior and grows caudad.

find differences in the maxillary dentition, which might suggest the presence of two maxillary dental laminae in additional lineages. Our results show that there is indeed a consistent difference in anterior versus posterior tooth morphologies in other advanced snake lineages (Supplementary Fig. 5d–x, Supplementary Table 3). In contrast, the maxillary teeth of the examined boids do not show a morphological difference between anterior and posterior teeth (Supplementary Fig. 5f). These results suggest the possible presence of two dental laminae in other non-front-fanged advanced snake lineages, and provide additional support to our proposed model.

The developmental uncoupling of the posterior from the anterior tooth region could have allowed the posterior teeth to evolve independently and in close association with the venom gland. Subsequently, the posterior teeth and venom gland could have become modified and formed the fang-gland complex—an event that underlies the massive radiation of advanced snakes during the Cenozoic era<sup>6,14,15</sup>.

## METHODS SUMMARY

**Snake embryos.** Snake eggs and embryos were acquired in accordance with local and international regulations from European, Israeli and Australian breeders and zoos. Eggs were incubated at 30 °C and embryos fixed in 4% paraformaldehyde in PBS at 4 °C overnight. They were dehydrated through graded methanols and stored at –18 °C.

**In situ hybridization.** The RNA probe was based on the partial PCR product of sonic hedgehog using the complementary DNA of a one-day-old rhombic night adder (*Causus rhombeatus*) embryo as template. Hybridization was performed according to standard protocols. In all species examined, the odontogenic band (tooth-forming epithelium) always expressed *shh* (Fig. 2, Supplementary Fig. 2). This shows that, as in other vertebrate groups<sup>10–13</sup>, *shh* is also a marker for odontogenic epithelium in snakes. A list of embryos studied can be found in Supplementary Table 1.

**Histology.** Embryos were dehydrated using ethanol or methanol, cleared in HistoClear or tetrahydronaphthalene and embedded in paraffin. Sections were cut at 5–7 µm and stained with H&E or counterstained using neutral red. A list of embryos studied can be found in Supplementary Table 2.

**Three-dimensional modelling.** Schematic three-dimensional models were drawn from analyses of the serial sections of the embryos using a Nikon Eclipse E800 microscope. All models were drawn using Adobe Illustrator and Adobe Photoshop.

**Scanning electron microscopy.** The maxillary bone on one side was dissected out of the specimen, allowed to dry and mounted on a stub, using double-sided tape, with teeth pointing upwards. Specimens were sputter-coated with gold, and examined using a JEOL JSM-T300 scanning electron microscope, at an acceleration voltage of 15 kV. A list of specimens examined, with their museum numbers, can be found in Supplementary Table 3.

**Ablation experiment.** The ablation experiment was performed as previously described<sup>29</sup>.

Received 2 November 2007; accepted 19 June 2008.

- Kochva, E. The origin of snakes and evolution of the venom apparatus. *Toxicon* **25**, 65–106 (1987).
- Kochva, E. in *Biology of the Reptilia* Vol. 8 (eds Gans, C. & Gans, K. A.) 43–162 (Academic Press, 1978).
- Jackson, K. The evolution of venom-delivery systems in snakes. *Zool. J. Linn. Soc.* **137**, 337–354 (2003).
- Kardong, K. V. Protovipers and the evolution of snake fangs. *Evolution* **33**, 433–443 (1979).
- Jackson, K. The evolution of venom-conducting fangs: insights from developmental biology. *Toxicon* **49**, 975–981 (2007).
- Vidal, N. Colubroid systematics: Evidence for an early appearance of the venom apparatus followed by extensive evolutionary tinkering. *J. Toxicol. Toxin Rev.* **21**, 21–41 (2002).
- Jackson, K. & Fritts, T. H. Evidence from tooth surface morphology for a posterior maxillary origin of the proteroglyph fang. *Amphibia-Reptilia* **16**, 273–288 (1995).
- Kardong, K. V. Evolutionary patterns in advanced snakes. *Am. Zool.* **20**, 269–282 (1980).
- Kardong, K. V. The evolution of the venom apparatus in snakes from colubrids to vipers and elapids. *Mem. Inst. Butantan* **46**, 105–118 (1982).
- Cobourne, M. T. & Sharpe, P. T. Sonic hedgehog signaling and the developing tooth. *Curr. Top. Dev. Biol.* **65**, 255–287 (2004).
- Fraser, G. J., Berkovitz, B. K., Graham, A. & Smith, M. M. Gene deployment for tooth replacement in the rainbow trout (*Oncorhynchus mykiss*): a developmental model for evolution of the osteichthyan dentition. *Evol. Dev.* **8**, 446–457 (2006).

- Fraser, G. J., Graham, A. & Smith, M. M. Developmental and evolutionary origins of the vertebrate dentition: molecular controls for spatio-temporal organisation of tooth sites in osteichthyans. *J. Exp. Zool. B* **306B**, 183–203 (2006).
- Yamanaka, A., Yasui, K., Sonomura, T. & Uemura, M. Development of heterodont dentition in house shrew (*Suncus murinus*). *Eur. J. Oral Sci.* **115**, 433–440 (2007).
- Kuch, U., Muller, J., Modden, C. & Mebs, D. Snake fangs from the Lower Miocene of Germany: Evolutionary stability of perfect weapons. *Naturwissenschaften* **93**, 84–87 (2006).
- Fry, B. G. et al. Evolution of an arsenal: Structural and functional diversification of the venom system in the advanced snakes (Caenophidia). *Mol. Cell. Proteomics* **7**, 215–246 (2007).
- Vidal, N. et al. The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. *C. R. Biol.* **330**, 182–187 (2007).
- Young, B. & Kardong, K. Dentitional surface features in snakes (Reptilia: Serpentes). *Amphibia-Reptilia* **17**, 261–276 (1996).
- Frazzetta, T. H. Studies on the morphology and function of the skull in the Boidae (Serpentes). Part II. Morphology and function of the jaw apparatus in *Python sebae* and *Python molurus*. *J. Morphol.* **118**, 217–296 (1966).
- Bogert, C. M. Dentitional phenomena in cobras and other elapids with notes on adaptive modifications of fangs. *Bull. Am. Mus. Nat. Hist.* **131**, 285–357 (1943).
- Cope, E. D. in *US National Museum Annual Report for 1898* 153–1270 (US Govt Print. Off., 1900).
- Boulenger, G. A. Remarks on the dentition of snakes and on the evolution of the poison-fangs. *Proc. Zool. Soc. Lond.* **40**, 614–616 (1896).
- Anthony, J. Essai sur l'évolution anatomique de l'appareil venimeux des Ophidiens. *Ann. Sci. Nat. Zool.* **17**, 7–53 (1955).
- Shayer-Wollberg, M. & Kochva, E. Embryonic development of the venom apparatus in *Causus rhombeatus* (Viperidae, Ophidia). *Herpetologica* **23**, 249–259 (1967).
- Kochva, E. Development of the venom gland and the trigeminal muscles in *Vipera palaestinae*. *Acta Anat.* **52**, 49–89 (1963).
- Odadia, M. Embryonic development of Duvernoy's gland in the snake *Natrix tessellata* (Colubridae). *Copeia* **1984**, 516–521 (1984).
- Gygax, P. Entwicklung, Bau und Funktion der Giftdrüse (Duvernoy's gland) von *Natrix tessellata*. *Acta Trop.* **28**, 226–274 (1971).
- Buchtova, M. et al. Embryonic development of *Python sebae* - II: Craniofacial microscopic anatomy, cell proliferation and apoptosis. *Zoology* **110**, 231–251 (2007).
- Reshef, R. *Ontogenetic Mechanisms of Teeth and Oral Gland Development in Snakes - A Possible Contribution to the Understanding of Evolutionary Processes in Reptiles*. PhD thesis, Univ. Tel-Aviv (1994).
- Reshef, R., Odadia, M., Wollberg, M. & Kochva, E. Snake yolk sac as a site for *in vivo* organ incubation: a new method in the research of snake embryo development. *J. Exp. Zool.* **270**, 538–546 (1994).
- Martin, H. Recherches sur le développement de l'appareil venimeux de la *Vipera aspis*. *C. R. Assoc. Anat.* **1**, 56–66 (1899).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the following persons and institutions who helped us or contributed material used in this study: J. W. Arntzen, M. Brittiijn, M. de Boer, R. van Deutekom, N. Dunstan, K. Van Egmond, P. L. Y. Fung, I. Gavrilov, W. Getreuer, J. Hanken, E. Heida, I. Ineich, T. de Jong, K. V. Kardong, M. Lautenbach, J. Losos, D. Millar, C. Pepermans, J. M. Richman, J. Rosado, R. de Ruiters, P. Schilperoord, M. M. Smith, S. Soubzmaigne, N. Vidal, E. M. Wielhouwer, J. Wolterling, National Museum of Natural History Naturalis Leiden, Reptilezoo Serpo, Muséum National d'Histoire Naturelle Paris, AQIS, DEH, APCG and DWLBC (Australia). This work received funding from the following sources: a Toptalent grant from the Netherlands Organization for Scientific Research (NWO; F.J.V.), a Smart Mix grant from the Dutch government (M.K.R.), a Valorisation grant from the Dutch Technology Foundation (STW; M.K.R., F.J.V., B.G.F.), the Curatoren fund (F.J.V.), the LUSTRA fund (F.J.V.), the Australian Research Council and the Australian Academy of Science (B.G.F.), DEST-ISL (B.G.F.), Whitman College (K.J.), a NWO visitors grant (M.K.R., B.G.F.) and the Leiden University Fund (F.J.V.).

**Author Contributions** F.J.V.: study concept and design; embryo and skull collection, acquisition and processing; probe synthesis; *in situ* hybridizations; histology; figures; paper. J.F.A.: three-dimensional reconstructions, figures. K.J.: scanning electron microscopy. R.R.: study concept, ablation experiment, histology. E.K.: study concept, embryological data. K.V., I.v.d.B., M.v.A.: *in situ* hybridizations, histology. M.A.G.d.B.: probe design, *in situ* hybridizations. A.B.: *Naja* embryo collection. P.J.M.: supply of *Naja* material. B.G.F.: study concept, supply of *Causus* material. A.W.: provision of laboratory space. E.B., F.W.: morphometric analyses. M.K.R. (project leader): study concept and design, provision of funding and laboratory space.

**Author Information** Sonic hedgehog complementary DNA clone sequence for the rhombic night adder, *Causus rhombeatus*, has been deposited in the Genbank database under accession number EU236145. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.K.R. ([m.k.richardson@biology.leidenuniv.nl](mailto:m.k.richardson@biology.leidenuniv.nl)).

## LETTERS

# The abundance threshold for plague as a critical percolation phenomenon

S. Davis<sup>1</sup>, P. Trapman<sup>2</sup>, H. Leirs<sup>3,4</sup>, M. Begon<sup>5</sup> & J. A. P. Heesterbeek<sup>1</sup>

Percolation theory is most commonly associated with the slow flow of liquid through a porous medium, with applications to the physical sciences<sup>1</sup>. Epidemiological applications have been anticipated for disease systems where the host is a plant or volume of soil<sup>2,3</sup>, and hence is fixed in space. However, no natural examples have been reported. The central question of interest in percolation theory<sup>4</sup>, the possibility of an infinite connected cluster, corresponds in infectious disease to a positive probability of an epidemic. Archived records of plague (infection with *Yersinia pestis*) in populations of great gerbils (*Rhombomys opimus*) in Kazakhstan have been used to show that epizootics only occur when more than about 0.33 of the burrow systems built by the host are occupied by family groups<sup>5</sup>. The underlying mechanism for this abundance threshold is unknown. Here we present evidence that it is a percolation threshold, which arises from the difference in scale between the movements that transport infectious fleas between family groups and the vast size of contiguous landscapes colonized by gerbils. Conventional theory predicts that abundance thresholds for the spread of infectious disease arise when transmission between hosts is density dependent such that the basic reproduction number ( $R_0$ ) increases with abundance, attaining 1 at the threshold. Percolation thresholds, however, are separate, spatially explicit thresholds that indicate long-range connectivity in a system and do not coincide with  $R_0 = 1$ . Abundance thresholds are the theoretical basis for attempts to manage infectious disease by reducing the abundance of susceptibles, including vaccination and the culling of wildlife<sup>6–8</sup>. This first natural example of a percolation threshold in a disease system invites a re-appraisal of other invasion thresholds, such as those for epidemic viral infections in African lions (*Panthera leo*), and of other disease systems such as bovine tuberculosis (caused by *Mycobacterium bovis*) in badgers (*Meles meles*).

In infectious disease epidemiology, there is an emphasis on thresholds that can be traced back to the threshold property of the basic reproduction number,  $R_0$ , defined here as the expected number of new infections arising from a single infectious host in a population of susceptibles. Values less than 1 imply that infectious individuals fail to replace themselves and that the disease inevitably dies out, whereas values greater than 1 imply that invasion of the disease into the host population is possible<sup>9</sup>. The difficulties with using reproduction numbers such as  $R_0$  when a host population is spatially or socially structured are generally understood<sup>10,11</sup>. Nevertheless, they remain the theoretical context for interpreting empirically observed thresholds<sup>12</sup> including the abundance threshold for plague epizootics<sup>5,13</sup>. Here, though, we present evidence that the plague threshold is a natural percolation phenomenon, and cannot be interpreted as empirical support for conventional thresholds based on  $R_0$ .

To explain this, we begin with a sketch of percolation theory. We then describe aspects of the population biology of great gerbils, and the landscapes they inhabit, which suggest percolation theory is an appropriate approach to understanding plague epizootics. A network model is then elaborated. The purpose of this model is to bring together three spatial scales: (i) that of flea movements responsible for plague transmission between family groups of great gerbils; (ii) the dimensions of the contiguous landscapes inhabited by great gerbils; and (iii) the scale at which plague monitoring is conducted in Kazakhstan. We argue that the network model shows how the three spatial scales are together responsible for the empirical observation of an abrupt threshold for plague.

Percolation theory concerns the behaviour of connected clusters in random networks<sup>4</sup>. Whether an infectious disease will spread among a population of hosts that have a fixed position in space and may only infect their nearest neighbours is well recognized as a percolation problem<sup>14,15</sup>. Equally recognized is the relevance of percolation theory to epidemics on networks in general, and lattice models in particular<sup>16–18</sup>. Among empiricists, the theory has been used to postulate the existence of spatial thresholds for the spread of fungal parasites of plants<sup>2</sup>.

A simple case of percolation is when bonds form independently between adjacent points on the plane square lattice with probability  $p$ . In network terminology, each point on the lattice is a vertex and if a bond is present there is said to be an open edge between the two vertices. A 'cluster' is then defined as a set of vertices connected by open edges. In percolation theory the questions of interest are whether an infinite cluster exists, whether this cluster is unique and whether a given vertex will belong to such a cluster (the existence of an infinite cluster does not imply that all vertices belong to it). The pertinent result is that there is a critical value of the probability  $p$ , denoted by  $p_c$ , below which every vertex belongs to a small (finite) cluster. For  $p > p_c$ , an infinite cluster exists and there is a (strictly) positive probability that a given vertex belongs to it. In the case of bond percolation on the plane square lattice, the infinite cluster is unique and  $p_c$  is equal to  $1/2$  (ref. 19). For numerous other networks, such a  $p_c$  also exists, though its value depends on the geometry of the network. In epidemiological applications, the given vertex corresponds to a primary case of an infectious disease and then  $p_c$  is an invasion threshold because the size of the cluster that the primary case belongs to corresponds to the final size of the epidemic. That is, when  $p < p_c$  the spread from the primary case is limited to a small cluster of hosts.

Great gerbils are desert rodents inhabiting vast areas of central Asia. They form family groups and rely on underground burrow systems to help protect them from predators and the extreme temperatures. The burrow systems vary in size but are often large, complex constructions

<sup>1</sup>Theoretical Epidemiology, Faculty of Veterinary Medicine, University of Utrecht, Yalelaan 7, 3584 CL Utrecht, The Netherlands. <sup>2</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, PO Box 85500, 3508 GA Utrecht, The Netherlands. <sup>3</sup>Department of Biology, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium. <sup>4</sup>Danish Pest Infestation Laboratory, University of Aarhus, Faculty of Agricultural Sciences, Department of Integrated Pest Management, Skovbrynet 14, DK-2800 Kongens Lyngby, Denmark. <sup>5</sup>Host-Parasite Biology Research Group, School of Biological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK.



representing the efforts of many generations. The vegetation above and around the burrow systems disappears such that a disc of bare earth forms, typically 30 m in diameter. These discs, and the patterned landscapes they create, are highly visible on satellite images (Fig. 1). Such images bring into clarity the regular spatial arrangement of burrow systems and the extent of the landscapes inhabited by great gerbils. The abundance of great gerbils, expressed as variation in the proportion of burrow systems inhabited, fluctuates over time<sup>20</sup>. More often than not, less than half of the burrow systems are occupied by family groups.

Plague bacteria are transmitted between great gerbils by fleas (mostly of the genus *Xenopsylla*) that inhabit the burrow systems. So an infectious flea has easy access to great gerbils living in the same burrow system but relatively limited access to those in surrounding burrow systems. Classifying whole family groups as either susceptible, infectious or recovered has been shown empirically to be a meaningful way of describing the distribution of plague in a great gerbil population<sup>21</sup>. Hence, in the network model we now present, vertices represent family groups rather than individual great gerbils. The random networks are generated from a landscape of burrow systems having the same structural characteristics as those observed on satellite images: 2.05 burrow systems per hectare with a buffer of 30 m (see Supplementary Information). Networks are formed from this landscape by choosing (random) subsets of the burrow systems to be vertices occupied by family groups. The size of the random subset is determined by the abundance of the host population, measured as occupancy.

Transmission of plague between family groups relies on the transport of infectious fleas between burrow systems. It occurs when an infectious flea migrates to its burrow system entrance and successfully

jumps to a passing gerbil. In the network model, then, a transmission event is when an infectious flea attempts to leave its burrow system. We assume a flea cannot differentiate between a gerbil that belongs to the burrow system it is trying to leave and a gerbil from another burrow system, so there is a chance the flea fails to disperse. Transmission events only occur at infected vertices and hence occur at a rate proportional to the number of infected vertices. If the vertex that receives the infectious flea is susceptible then it becomes infected. To determine the vertex that received the infectious flea, we used gerbil movement data recorded during mark–recapture studies<sup>22</sup>, combined with older field studies of flea dispersal in which fleas were marked using radionucleotides and their movements observed directly<sup>23,24</sup>, to construct a set of weights for the vertices surrounding the infected vertex, including the infected vertex itself (see Supplementary Information). Most (more than 95%) observed flea movements, and great gerbil movements, were less than 200 m.

Each sample of great gerbils, tested for plague, is linked to a ‘sector’ (10 km × 10 km area) in the archives. The PreBalkhash focus consists of approximately 352 such sectors, only a fraction of which are visited each surveillance season. The results of the network model at this scale are shown in Fig. 2. Of principal interest is how the probability of an epizootic varies with the abundance of great gerbils expressed as the proportion of burrow systems occupied (varied between 0.01 and 0.49). In the model, there are three possible outcomes when a single family group is infected: (i) the infection dies out within this first family group, failing to spread to any other; (ii) infection spreads but only within a small cluster of family groups (there is a minor outbreak); or (iii) infection spreads within a large cluster (epizootic). A percolation threshold emerges provided the criterion defining a large cluster is that plague spreads more than several kilometres from the



**Figure 1 | The regular, star-like pattern created by burrow systems, visible on satellite images.** Patches of bare earth form above and around the burrow systems dug by great gerbils and strongly reflect the sunlight. Each

bright disc represents a burrow system 10–40 m in diameter. The image was captured using the publicly available software Google Earth (<http://earth.google.com/>). Copyright 2008 DigitalGlobe; Europa Technologies.

site of first infection. As the distance used to define a large cluster is increased, the curves in Fig. 2 approach a non-differentiable curve similar to that found for percolation on infinite lattices<sup>4</sup>; that is, there is an interval of  $p$  over which the chance of a large-scale plague epizootic increases abruptly as  $p$  increases. When epizootics are defined as spread of plague over shorter distances, the chance of an epizootic increases steadily rather than abruptly. The spatial scale at which an epizootic is defined to have occurred therefore plays a large role in whether results indicate an abrupt threshold.

The percolation threshold in the network model occurs at about 0.31. This may be compared to the abundance threshold directly estimated from the plague archives by noting that the estimated threshold in a single-abundance model<sup>3</sup>, which performed almost as well as the best, dual-abundance model (a difference in Akaike's information criterion of just 0.71), was 0.33 (95% confidence interval: 0.287, 0.373). That is, the percolation threshold falls within the 95% confidence interval of the threshold estimated from the plague archives. Hence we may, notwithstanding that some parameters in the network model were set to orders of magnitude, at least conclude that there is certainly no conflict between the quantitative results

from the network model and the empirical abundance threshold, previously estimated. The sensitivity of the model to its parameters, including changes to the frequency of long-range flea movements, reveals a robust percolation result (see Supplementary Information).

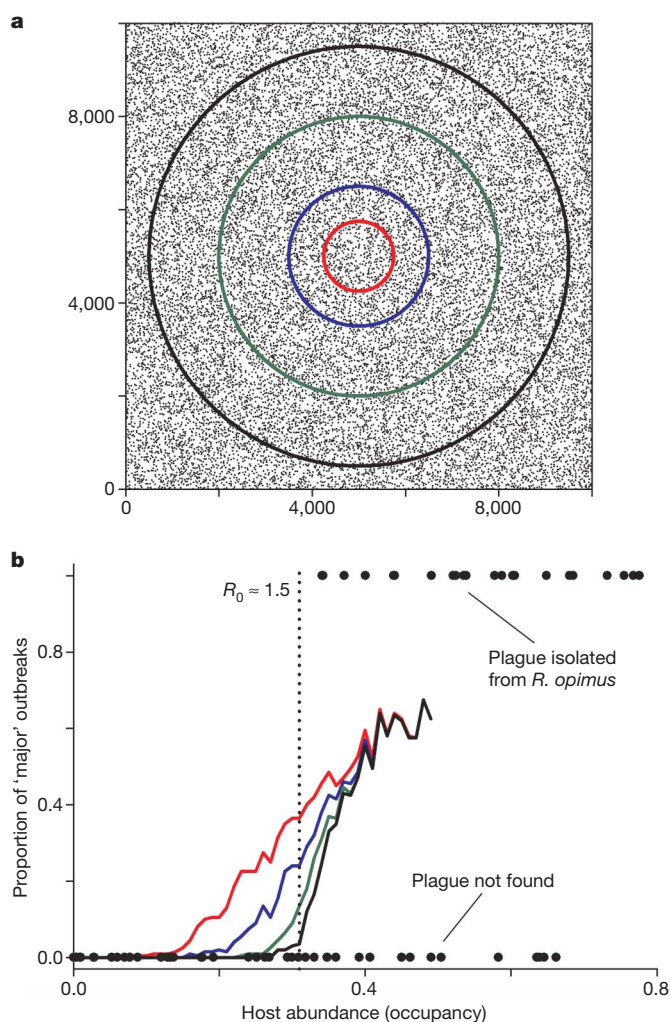
A percolation threshold in an infectious-disease system takes into account the local depletion of susceptibles that arises from spatial restrictions on who can (or is likely to) infect whom. It is the point at which the landscape is sufficiently filled with hosts or host groups for the disease to continue to spread despite this barrier. This is distinctly different from the conventional understanding of invasion thresholds based on  $R_0$  and models of randomly mixing populations. We measured  $R_0$  directly from the network model simulations by recording the number of vertices infected by the first infected vertex and then averaging over many simulations (consistent with the definition of  $R_0$  as the expected number of secondary cases resulting from a single primary case in a fully susceptible population). At an occupancy of 0.31, the value of occupancy at which the percolation threshold occurs,  $R_0$  is around 1.5. More generally, for epidemics on lattice-like networks, the spatial constraints on who can infect whom mean that the expected number of secondary cases can be well above 1 but still only small outbreaks are possible<sup>25</sup>. How much larger than 1  $R_0$  must be for sustained spread to occur will depend on the details of particular systems, including the geometry of the spatial arrangement of hosts.

Our conclusion is that epizootics of sylvatic plague are a natural percolation phenomenon and that this is the mechanism (together with the spatial scale of surveillance in Kazakhstan) responsible for the abrupt population threshold seen in the plague archives. The empirical basis for this statement is the contrast between the lengths of movements of fleas between burrow systems and the scale of contiguous landscapes inhabited by the host. As the first natural example of a percolation threshold for disease invasion to be reported, it is even more interesting because the hosts are vertebrates. One consequence is a re-appraisal of other invasion thresholds such as the minimum densities of susceptible African lions for epidemic viral infections in the Serengeti<sup>26</sup>. The immune (or nearly immune) prides may create a network of hosts that is difficult for the aetiological agent to percolate through. This depends critically on whether transmission of the viruses between prides of lions occurs by (aggressive) contacts between lions from different prides, or whether mobile reservoir hosts such as spotted hyenas or nomadic lions are acting as a vector<sup>27</sup>. A second system that may warrant re-evaluation is bovine tuberculosis in badger populations, wherein badger setts may be analogous to great gerbil burrow systems. If percolation is the pertinent perspective, such that spread of tuberculosis in a badger population is dependent on long-range connectivity, then any lines of reasoning about the ability of badgers to sustain the disease (as a reservoir independent of transmission back and forth within the cattle population) must also be seen in that light.

## METHODS SUMMARY

Random networks of vertices representing occupied burrow systems were generated from a single landscape of burrow systems, the construction of which was based on the spatial arrangement of burrow systems observed on satellite images. Epizootics, or failed epizootics, were initiated on these networks by choosing the burrow system closest to the centre of the landscape to be infected. The spread of plague from this burrow system (vertex) was simulated by using a state vector and an adjacency matrix. The state vector had length equal to the number of occupied burrow systems (vertices), and each entry was either 0, 1 or 2, these values respectively indicating a susceptible, infected or recovered family group. Epizootics on the random networks essentially consisted of events occurring at infected vertices (the dispersal of an infected flea or the recovery of a family group from plague).

Time was modelled explicitly by defining the gap between events as a random variable with an exponential distribution and a rate proportional to the number of infected vertices. The consequences of events were dealt with as they occurred: that is, for each event, a random infected vertex was chosen at which the event happened and if a flea dispersal event occurred then which burrow system



**Figure 2 | The results of the network model for plague epizootics in great gerbils.** The landscape (a) is a 10 km × 10 km area, and the results (b) are expressed as the fraction of outbreaks that give rise to new infections at least 750 m, 1.5 km, 3 km and 4.5 km from the site of initial infection, shown as solid lines coloured red, blue, green and black, respectively, and corresponding to the spread of plague beyond the circles of the same colour in a. At the estimated percolation threshold,  $R_0$  does not coincide with 1. Empirical observations (filled circles) of epizootics show an abrupt threshold at about 0.33 (ref. 5).



received the flea was also decided. If the burrow system that received the flea was susceptible then it immediately became infected (that is, the entry in the state vector changed from 0 to 1). If the event was a recovery event then the entry in the state vector changed from 1 to 2. A simulation ended either when there were no 1 entries in the state vector or if the distance between a newly infected vertex and the initially infected vertex exceeded 4.5 km.

Received 31 January; accepted 30 April 2008.

- Sahimi, M. *Applications of Percolation Theory* (Taylor & Francis, London, 1994).
- Otten, W., Bailey, D. J. & Gilligan, C. A. Empirical evidence of spatial thresholds to control invasion of fungal parasites and saprotrophs. *New Phytol.* **163**, 125–132 (2004).
- Bailey, D. J., Otten, W. & Gilligan, C. A. Saprotrophic invasion by the soil-borne fungal plant pathogen *Rhizoctonia solani* and percolation thresholds. *New Phytol.* **146**, 535–544 (2000).
- Grimmett, G. *Percolation* (Springer-Verlag, Berlin, 1999).
- Davis, S. *et al.* Predictive thresholds for plague in Kazakhstan. *Science* **304**, 736–738 (2004).
- Artois, M., Delahay, R., Guberti, V. & Cheeseman, C. Control of infectious diseases of wildlife in Europe. *Vet. J.* **162**, 141–152 (2001).
- Caley, P., Hickling, G. J., Cowan, P. E. & Pfeiffer, D. U. Effects of sustained control of brushtail possums on levels of *Mycobacterium bovis* infection in cattle and brushtail possum populations from Hohotaka, New Zealand. *NZ Vet. J.* **47**, 133–142 (1999).
- Barlow, N. D. The ecology of wildlife disease control: simple models revisited. *J. Appl. Ecol.* **33**, 303–314 (1996).
- Diekmann, O. & Heesterbeek, J. A. P. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation* (John Wiley, Chichester, UK, 2000).
- Keeling, M. J. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B* **266**, 859–867 (1999).
- Trapman, P. *On Stochastic Models for the Spread of Infections*. PhD thesis, Vrije Univ. Amsterdam (2006).
- Lloyd-Smith, J. O. *et al.* Should we expect population thresholds for wildlife disease? *Trends Ecol. Evol.* **20**, 511–519 (2005).
- Davis, S. *et al.* Empirical assessment of a threshold model for sylvatic plague. *J. R. Soc. Interface* **4**, 649–657 (2007).
- Grassberger, P. On the critical behaviour of the general epidemic process and dynamical percolation. *Math. Biosci.* **63**, 157–172 (1983).
- Mollison, D. Spatial contact models for ecological and epidemic spread. *J. R. Statist. Soc. B* **39**, 283–326 (1977).
- Keeling, M. J. & Eames, T. D. Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307 (2005).
- Newman, M. E. J. The structure and function of complex networks. *Siam Rev.* **45**, 167–256 (2003).
- Trapman, P. On analytical approaches to epidemics on networks. *Theor. Popul. Biol.* **71**, 160–173 (2007).
- Kesten, H. The critical probability of bond percolation on the square lattice equals  $1/2$ . *Commun. Math. Phys.* **74**, 41–59 (1980).
- Kausrud, K. L. *et al.* Climatically driven synchrony of gerbil populations allows large-scale plague outbreaks. *Proc. R. Soc. B* **274**, 1963–1969 (2007).
- Davis, S. *et al.* Plague metapopulation dynamics in a natural reservoir: the burrow-system as the unit of study. *Epidemiol. Infect.* **135**, 740–748 (2007).
- Begon, M. *et al.* Epizootiological parameters for plague (*Yersinia pestis* infection) in a natural reservoir in Kazakhstan. *Emerg. Infect. Dis.* **12**, 268–273 (2006).
- Rudenchik, Yu. V., Soldatkin, I. S., Severova, E. A., Mokiyevech, N. A. & Klimova, Z. I. Quantitative evaluation of the possibility of the territorial advance of epizooty of plague in the population of the *Rhombymus opimus* (northern Karakum). *Zool. Zh.* **46**, 117–123 (1967).
- Korneyev, G. A. *The Quantitative Characteristics of Parasitic Exchange in Some Mammal Species Resulting from Simulation of Epizootics in Desert Biocenoses* (Microbe Institute, Saratov, 1968).
- Mollison, D. The dependence of epidemic and population velocities on basic parameters. *Math. Biosci.* **107**, 255–287 (1991).
- Packer, C. *et al.* Viruses of the Serengeti: patterns of infection and mortality in African lions. *J. Anim. Ecol.* **68**, 1161–1178 (1999).
- Roelke-Parker, M. *et al.* A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature* **379**, 441–445 (1996).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank our colleagues V. Dubyanski and V. Ageyev for their assistance with interpreting publications and discussions about the PreBalkhash plague system, its surveillance and the host population of great gerbils. We thank J. Taxis for help with coding the network model. We also thank J. Rees for her advice and encouragement. The work of S.D. and J.A.P.H. was supported by the Netherlands Organisation for Scientific Research (NWO/ZonMw grant 918.56.620). We also acknowledge the support of the UK Joint Environment and Human Health Programme (funders: Natural Environment Research Council, Defra, Environment Agency, Ministry of Defence, Medical Research Council).

**Author Contributions** S.D. proposed and pursued the question of what mechanism generated the plague threshold, constructed and analysed the network model and wrote the paper. P.T. introduced S.D. to percolation theory and contributed to the writing of the Supplementary Information. All authors discussed percolation and the plague system and made comments on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.D. (S.A.Davis@uu.nl).



## LETTERS

# An Fgf/Gremlin inhibitory feedback loop triggers termination of limb bud outgrowth

Jamie M. Verheyden<sup>1</sup> & Xin Sun<sup>1</sup>

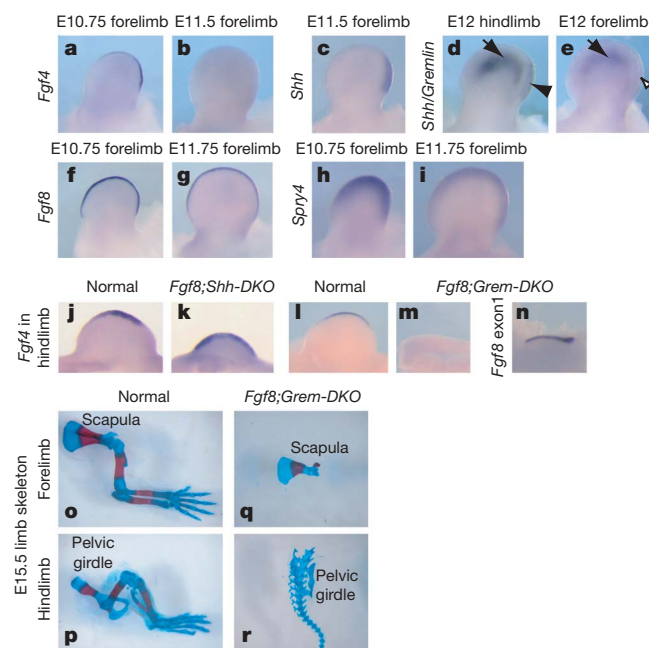
During organ formation and regeneration a proper balance between promoting and restricting growth is critical to achieve stereotypical size. Limb bud outgrowth is driven by signals in a positive feedback loop involving fibroblast growth factor (Fgf) genes, *sonic hedgehog* (*Shh*) and *Gremlin1* (*Grem1*)<sup>1</sup>. Precise termination of these signals is essential to restrict limb bud size<sup>2–4</sup>. The current model predicts a sequence of signal termination consistent with that in chick limb buds<sup>4</sup>. Our finding that the sequence in mouse limb buds is different led us to explore alternative mechanisms. Here we show, by analysing compound mouse mutants defective in genes comprising the positive loop, genetic evidence that FGF signalling can repress *Grem1* expression, revealing a novel Fgf/*Grem1* inhibitory loop. This repression occurs both in mouse and chick limb buds, and is dependent on high FGF activity. These data support a mechanism where the positive Fgf/*Shh* loop drives outgrowth and an increase in FGF signalling, which triggers the Fgf/*Grem1* inhibitory loop. The inhibitory loop then operates to terminate outgrowth signals in the order observed in either mouse or chick limb buds. Our study unveils the concept of a self-promoting and self-terminating circuit that may be used to attain proper tissue size in a broad spectrum of developmental and regenerative settings.

Several models have recently been formulated to explain the control of appendage size<sup>5–8</sup>. The models focus on how a signal in constant supply is translated into a threshold of growth capability. In contrast, evidence from vertebrate limb development suggests that precise termination of growth signals is a key mechanism that restricts limb bud size<sup>2–4</sup>. These signals include Fgfs (*Fgf4*, *Fgf8*, *Fgf9* and *Fgf17*) expressed in the apical ectodermal ridge (AER and AER-Fgfs), and *Shh* and *Grem1* expressed in the underlying mesenchyme. They function in a transcriptional feedback loop (Fgf/*Shh* loop) to induce and sustain each other's expression<sup>1,9–11</sup>.

The current model for breakdown of the Fgf/*Shh* loop is based on the observation that current and former *Shh*-expressing cells (*Shh*-lineage cells) are unable to express *Grem1* in response to SHH induction<sup>4</sup>. Expansion of the *Shh*-lineage would lead to cessation of *Grem1* expression followed by that of *Fgf4* and then *Shh*. This sequence of signal termination is consistent with that observed in chick. However, we found that in mouse limb buds, *Fgf4* expression ceases first, followed by *Shh* and then *Grem1* (Fig. 1a–e). To identify alternative mechanisms of signal termination, we investigated the regulation of *Fgf4*, the first gene of the loop that ceases to be expressed in mouse limb buds. Although *Fgf4* itself is not essential for limb development<sup>9</sup>, it is regulated by essential genes, including *Shh* and *Grem1*<sup>10,12,13</sup>. Furthermore, termination of *Fgf4* expression coincides with a drop in collective AER-FGF activity<sup>14</sup> (Fig. 1f–i). Therefore the extinction of *Fgf4* expression serves as readout for the trigger that breaks down the Fgf/*Shh* loop.

*Fgf4* expression is severely reduced in *Shh* mutant and absent in *Grem1* mutant limb buds, but expanded and prolonged in *Fgf8*

AER-knockout (*Fgf8*-KO) forelimb buds<sup>10–13,15,16</sup>. We investigated whether these regulators act genetically upstream or downstream of each other to control *Fgf4* expression. To address if *Fgf8* represses *Fgf4* expression by inhibiting *Shh* maintenance of *Fgf4*, we inactivated both *Fgf8* and *Shh* in the limb buds by introducing a null allele of *Shh*<sup>17</sup> into the *Fgf8*-KO background<sup>15</sup> (*Msx2cre*; *Fgf8*<sup>del/fl</sup>; *Shh*<sup>−/−</sup> mutant, or *Fgf8*; *Shh*-DKO for double knockout). In *Fgf8*; *Shh*-DKO forelimb and hindlimb buds, *Fgf4* is detected in an expanded pattern



**Figure 1 | *Fgf8* repression of *Fgf4* expression is dependent on *Grem1* but not *Shh*.** a–n, Gene expression in mouse forelimb or hindlimb buds. a–i, In wild-type mouse limb buds, *Fgf4* expression terminates first, followed by *Shh* and then *Grem1*. In d and e, a combination of RNA probes is used to detect non-overlapping patterns of *Shh* (arrowhead) and *Grem1* (arrow) expression.

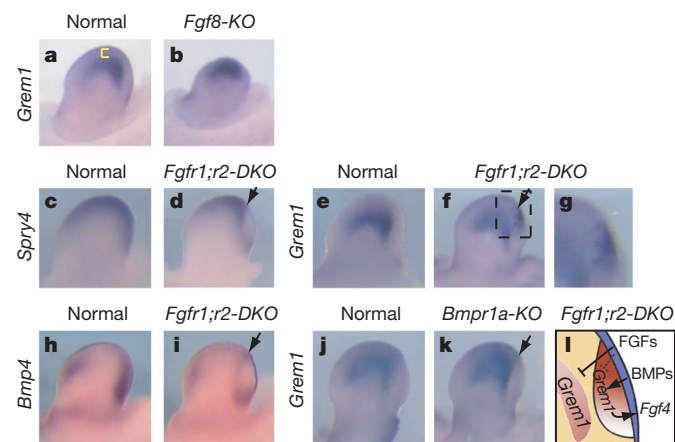
Both genes are expressed in the E12 hindlimb bud, which is at an earlier developmental stage than the E12 forelimb bud from the same embryo, where only *Grem1* is expressed ( $n = 4$ ). Downregulation of *Spry4* expression at E11.75 compared with E10.75 reflects decreased AER-FGF activity<sup>14</sup>, consistent with loss of *Fgf4* and reduced *Fgf8* expression. j–m, In E10.5 hindlimb buds, *Fgf4* expression is detected in the posterior two-thirds of the AER in normal hindlimb, expanded through the entire AER in the *Fgf8*; *Shh*-DKO mutant and absent in the *Fgf8*; *Grem1*-DKO mutant. n, Detection of the remaining exon 1 of the truncated *Fgf8* mRNA indicates that the AER is present. o–r, No forelimb or hindlimb elements are observed in *Fgf8*; *Grem1*-DKO skeletons.

<sup>1</sup>Laboratory of Genetics, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.

in the entire AER (Fig. 1j, k and data not shown), demonstrating that *Fgf8* repression of *Fgf4* expression is genetically downstream of *Shh*.

*Grem1* functions downstream of *Shh* to induce *Fgf4* expression<sup>10,11</sup>. To address if *Fgf8* represses *Fgf4* expression by inhibiting *Grem1*, we inactivated both *Fgf8* and *Grem1* in the limb buds by introducing a null allele of *Grem1* (ref. 13) into the *Fgf8*-KO background (*Msx2cre*; *Fgf8*<sup>del/β</sup>; *Grem1*<sup>-/-</sup> mutant, or *Fgf8*; *Grem1*-DKO). In *Fgf8*; *Grem1*-DKO limb buds, *Fgf4* is no longer maintained, even though the AER is present (Fig. 1l–n). With AER-*Fgf* expression severely compromised, all limb skeletal elements are absent (Fig. 1o–r), similar to the phenotype in *Fgf4* and *Fgf8* double-mutant limbs<sup>18</sup>. This loss of *Fgf4* expression in *Fgf8*; *Grem1*-DKO limb buds demonstrates that *Fgf8* repression of *Fgf4* is dependent on *Grem1*.

To understand the mechanism of this dependence, we investigated whether *Fgf8* represses *Grem1* expression. Consistent with this possibility, the *Grem1* domain is closer to the AER than normal in *Fgf8*-KO limb buds (Fig. 2a, b). As all AER-FGFs perform similar roles in limb bud outgrowth<sup>19</sup>, we tested a more general hypothesis that collective AER-FGF signalling could repress *Grem1* expression. In support of this, the *Grem1* expression domain is closer to the AER in various other *Fgf* and *Fgf* receptor (*Fgfr*) mutants (Supplementary Fig. 1). One caveat is that these mutant limb buds are smaller than normal, raising the possibility that the *Grem1* domain is closer to the AER because of reduction of the distal mesenchyme. To test FGF repression of *Grem1* more rigorously, we inactivated *Fgfr1* and *Fgfr2* in a small portion of the limb bud mesenchyme (*Shh*<sup>cre</sup>; *Fgfr1*<sup>co/co</sup>; *Fgfr2*<sup>c/c</sup> mutant, or *Fgfr1*; *r2*-DKO)<sup>20–22</sup>. We found that although FGF signalling is severely disrupted in *Fgfr*-inactivated cells, *Fgfr1*; *r2*-DKO limb buds exhibit normal size, shape and cell survival at E11.5 (Fig. 2c, d and data not shown). In this setting, *Grem1* is ectopically expressed within *Fgfr*-inactivated domain (Fig. 2e–g and Supplementary Fig. 2). Our loss-of-function data complement a previous observation that FGF-soaked beads can inhibit *Grem1* expression in chick limb buds<sup>23</sup>. These findings demonstrate that AER-FGF signalling is sufficient and necessary to repress *Grem1* expression in the distal mesenchyme.

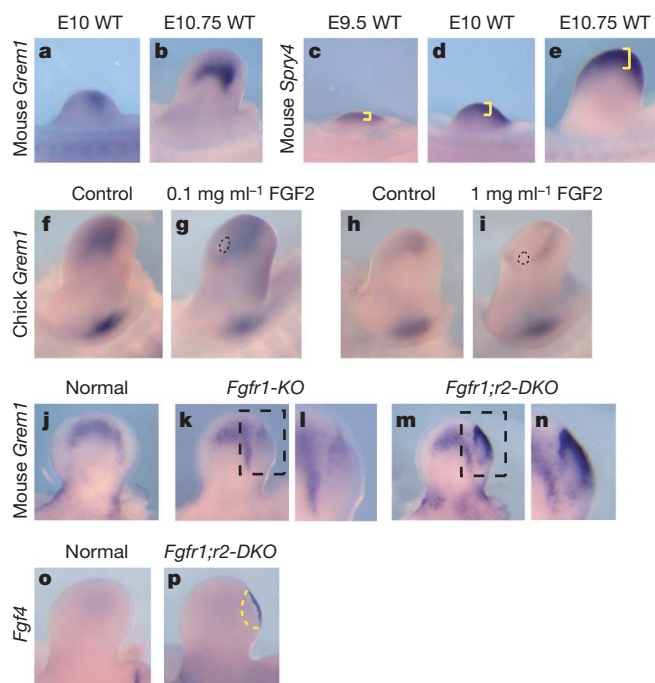


**Figure 2 | FGF signalling represses *Grem1* expression.** a–k, Gene expression in mouse forelimb buds at (a, b) E10.75, (c–i) E11 and (j, k) E11.5. a, The yellow bracket indicates distance between AER and high *Grem1* expression. c, d, Reduced *Spry4* expression in the posterior mesenchyme delineates *Fgfr*-inactivated domain. Arrows in d, f, i, k indicate anterior boundary of *Shh*<sup>cre</sup>-mediated receptor inactivation domain. e–g, *Grem1* is ectopically expressed in the distal portion of the *Fgfr*-inactivated domain. Limb buds shown in d and f are contralateral limb buds from the same embryo. Boxed region in f is magnified in g. h, i, *Bmp4* is reduced in *Fgfr*-inactivated domain, but is present in the overlying AER. j, k, No ectopic *Grem1* expression is detected in *Shh*<sup>cre</sup>; *Bmpr1a*<sup>fl/fl</sup> (*Bmpr1a*-KO) limb buds. l, Diagram depicting gene expression regulation within the *Fgfr*-inactivated domain in *Fgfr1*; *r2*-DKO limb buds as shown in g. BMPs from the AER may be required to promote ectopic *Grem1* (refs 23–27), leading to higher *Grem1* in the distal portion of the *Fgfr*-inactivated domain as shown in Figs 2g and 3l, n.

High levels of exogenous bone morphogenetic protein (BMP) have been shown to inhibit *Grem1* expression<sup>23–25</sup>. We found that *Bmp4* and *Bmp7* expression is reduced in the *Fgfr*-inactivated cells in *Fgfr1*; *r2*-DKO limb buds, raising the possibility that AER-FGFs repress *Grem1* by maintaining high BMP signalling (Fig. 2h, i and data not shown). However, inactivation of *Bmpr1a* with *Shh*<sup>cre</sup> does not lead to ectopic *Grem1* expression (Fig. 2j, k), suggesting that AER-FGF repression of *Grem1* is not mediated through BMPs. It remains possible that BMP signalling may be required to promote *Grem1* expression in parallel to FGF repression of *Grem1* (refs 23–27) (Fig. 2l).

To investigate the threshold requirement for FGF repression of *Grem1*, we compared *Grem1* expression to changes in FGF signalling. In mouse limb buds, downregulation of *Grem1* in the distal mesenchyme correlates with progressively higher levels of FGF signalling as development proceeds (Fig. 3a–e). This result is consistent with that observed in chick limb buds<sup>23,25</sup>. These gene-expression data led us to hypothesize that AER-FGF signalling represses *Grem1* in a dose-sensitive manner.

We tested this hypothesis in mouse and chick limb buds. In chick, implantation of beads soaked in 1 mg ml<sup>-1</sup> FGF2 leads to a clear repression of *Grem1* (*n* = 4/6, Fig. 3h, i), consistent with previous observations<sup>23</sup>. This repression is not observed using beads soaked in 0.1 mg ml<sup>-1</sup> FGF2 (*n* = 0/7, Fig. 3f, g), even though *Spry2* upregulation is detected adjacent to the beads, confirming FGF activity (data not shown). In mouse, ectopic *Grem1* expression is more intense in *Fgfr1*; *r2*-DKO limb buds than in *Shh*<sup>cre</sup>; *Fgfr1*<sup>co/co</sup> (*Fgfr1*-KO) limb buds (Fig. 3j–n). As there is less residual FGF signalling in *Fgfr1*; *r2*-DKO than in *Fgfr1*-KO limb buds (based on expression of FGF readouts,

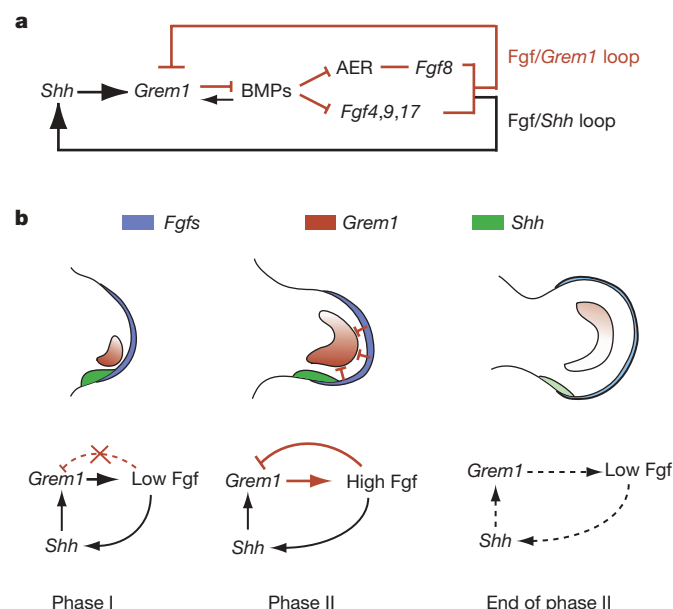


**Figure 3 | AER-FGF repression of *Grem1* expression is dose sensitive.** a–e, Correlation between *Grem1* repression in the distal mesenchyme and increased AER-FGF signalling (yellow brackets in c–e)<sup>14</sup>. f–i, Beads (circled) soaked in 1 mg ml<sup>-1</sup> FGF2 suppress *Grem1* expression distal to the bead, possibly working in combination with FGFs expressed from the AER (*n* = 4/6). No *Grem1* suppression is observed with 0.1 mg ml<sup>-1</sup> FGF2 (*n* = 7). j–n, Although ectopic *Grem1* expression is more intense in E11.5 *Fgfr1*; *r2*-DKO limb buds than in *Fgfr1*-DKO limb buds, *Grem1* expression outside the *Fgfr*-inactivated domain remains comparable. Boxed regions in k, m are magnified in l, n, respectively. o, p, Though absent in E11.75 normal limb bud, *Fgf4* expression persists in the posterior AER overlying the *Fgfr*-inactivated domain in *Fgfr1*; *r2*-DKO limb buds (delineated by yellow dashed line). WT, wild type.

data not shown), lower FGF signalling correlates with less *Grem1* repression. Thus data both from chick and mouse limb buds support the hypothesis that, during limb bud outgrowth, a progressive increase in AER-FGF level leads to increasing repression of *Grem1* in the distal mesenchyme.

To return to our question about the mechanism that abolishes *Fgf4* expression and triggers *Fgf/Shh* loop termination, we found that *Fgf4* expression is prolonged in *Fgfr1;2-DKO* forelimb buds at embryonic day (E)11.75 (Fig. 3o, p), probably as a result of ectopic *Grem1* expression<sup>11–13,25</sup>. These data demonstrate that FGF repression of *Grem1* plays a critical role in triggering the termination of limb-bud outgrowth signals.

The finding that AER-FGF signalling can repress *Grem1* expression reveals an inhibitory feedback loop (*Fgf/Grem1* loop) that is interconnected with the existing *Fgf/Shh* positive feedback loop (Fig. 4a). The dose dependency of this repression led us to propose a model (Fig. 4b) whereby positive and inhibitory feedback loops are coordinated first to promote (in phase I, with *Fgf/Shh* positive loop only) and later to terminate limb bud outgrowth (in phase II, with the induction of *Fgf/Grem1* inhibitory loop). In a wild-type limb bud in phase I (for example, at approximately E9.5–E10.5 in mouse forelimb bud, or stages 18–23 in chick wing bud), we hypothesize that AER-FGF concentration is too low to repress *Grem1* efficiently (Fig. 3c, d, g, i). Instead, AER-FGFs act through *Shh* and BMPs to upregulate *Grem1* (refs 10, 24, 25). As a result, *Grem1* is expressed in the distal mesenchyme abutting the AER (Fig. 3a) and efficiently promotes AER-FGF expression<sup>3,26,28,29</sup>. In phase I, the positive *Fgf/Shh* loop induces and sustains limb outgrowth signals, leading to a progressive increase in collective AER-FGFs entering phase II (Fig. 3c–e).



**Figure 4 | Model describing a self-promoting and self-terminating mechanism to control limb-bud outgrowth signals. a,** The inhibitory loop (outlined in red) in relation to the existing positive loop. Arrows indicate activation, whereas ‘T’-shaped lines indicate inhibition. BMP regulation of AER architecture indirectly affects *Fgf8* expression<sup>12,13</sup>. *Grem1* is also positively regulated by BMP signalling<sup>11,24,25,27</sup>. **b,** Model explaining how the two loops are used first to promote (phase I) and then terminate (phase II) signals. Dashed lines represent diminishing regulation whereas the dashed line with the cross through it emphasizes absence of regulation. In phase I, the positive regulatory loop operates to increase all signals. Transition to phase II occurs when AER-FGFs reach a level that confers efficient *Grem1* repression (represented by ‘T’-shaped lines in both distal and posterior mesenchyme). Together with mesenchymal growth, the *Grem1*-negative domain expands. Increasing distance between *Grem1*-expressing cells and *Fgf*- or *Shh*-expressing cells leads to inability of signals to maintain one another at the end of phase II.

We hypothesize that the transition to phase II occurs when AER-FGF signalling surpasses the threshold needed for *Grem1* repression in the distal mesenchyme, triggering the *Fgf/Grem1* inhibitory loop (Fig. 4b, for example, at approximately E10.5–E12 in mouse forelimb bud, or stages 23–27 in chick wing bud). This repression establishes a *Grem1*-negative domain separating *Grem1*-expressing cells and the AER (Fig. 2a). As development proceeds, the *Grem1*-negative domain expands both distally and posteriorly owing to mesenchymal growth. We postulate that this expansion would trigger different rate-limiting steps in mouse versus chick limb buds, leading to distinct sequences of signal termination. In a mouse limb bud, the size of the *Grem1*-negative domain would first exceed the distal range of *Grem1* protein diffusion, leading to downregulation of collective AER-FGFs followed by loss of *Shh* and then *Grem1* expression (Fig. 4b, end of phase II). Loss of *Grem1* expression would mark the beginning of AER degeneration and gradual extinction of *Fgf8* expression<sup>12,13</sup>. Conversely, in a chick limb bud, the size of the *Grem1*-negative domain would first exceed the anterior range of SHH diffusion, leading to loss of *Grem1* expression followed by extinction/reduction of different AER-FGFs and then termination of *Shh*. Thus, this model can explain the sequence of signal abrogation in both mouse and chick. We further postulate that in a wider spectrum of divergent species, parameters such as signal diffusion range, threshold requirement of signalling activity and extent of mesenchyme expansion dictate the timing of outgrowth signal termination.

There are two key differences between our model and the existing model of signal termination<sup>4</sup> (the *Shh*-lineage model). First, the *Shh*-lineage model only accounts for *Grem1* repression in posterior mesenchyme. Our model explains *Grem1* repression in both the posterior and distal mesenchyme, which accommodates the sequence of signal termination both in mouse and chick limb buds. Second, the molecular mechanisms at the core of the two models are distinct. In the *Shh*-lineage model, the factor responsible for cell-autonomous repression of *Grem1* in *Shh*-lineage cells has not been identified. Our finding that *Fgfr* inactivation allows *Grem1* expression in *Shh*-lineage cells (Figs 2f, g and 3k–n) suggests that maintenance of FGF signalling is essential for *Grem1* repression in this lineage. In our model, signal termination relies on FGF repression of *Grem1* expression. The finding that an FGF bead placed in the anterior chick limb bud downregulates *Grem1* expression<sup>23</sup> (Fig. 3i) indicates that FGF repression of *Grem1* can occur independently of the *Shh*-lineage influence.

In this study, we identified an inhibitory *Fgf/Grem1* feedback loop that operates both in mouse and chick limb buds. We propose a model whereby the known positive *Fgf/Shh* feedback loop acts to increase AER-FGF concentration, triggering the inhibitory loop, which in turn leads to extinction of outgrowth signals. These interconnected positive and inhibitory loops direct a limb outgrowth programme that, once initiated, can propagate and self-terminate.

## METHODS SUMMARY

**Generation of mutants.** *Fgf8;Shh-DKO* embryos were generated by mating *Msx2cre;Fgf8<sup>del/flox</sup>;Shh<sup>+/+</sup>* males to *Fgf8<sup>flox/flox</sup>;Shh<sup>+/+</sup>* females<sup>15,17</sup>. *Fgf8;Grem1-DKO* embryos were generated by crossing *Msx2cre;Fgf8<sup>del/flox</sup>;Grem1<sup>+/+</sup>* males to *Fgf8<sup>flox/flox</sup>;Grem1<sup>+/+</sup>* females<sup>13,15</sup>. *Fgfr1;2-DKO* embryos were generated by mating *Shh<sup>cre</sup>;Fgfr1<sup>co/co</sup>;Fgfr2<sup>+/+</sup>* males to *Fgfr1<sup>co/co</sup>;Fgfr2<sup>+/+</sup>* females<sup>20–22</sup>. *Bmpr1a-DKO* embryos were generated by mating *Shh<sup>cre</sup>;Bmpr1a<sup>fl/fl</sup>* males to *Bmpr1a<sup>fl/fl</sup>* females<sup>21,30</sup>.

**Bead implantation.** Beads were implanted in stage 22 limb buds. Gene expression was assayed after 12 h of incubation.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 28 January; accepted 13 May 2008.

Published online 25 June 2008.

- Niswander, L. Interplay between the molecular signals that control vertebrate limb development. *Int. J. Dev. Biol.* **46**, 877–881 (2002).
- Sanz-Ezquerro, J. J. & Tickle, C. Fgf signaling controls the number of phalanges and tip formation in developing digits. *Curr. Biol.* **13**, 1830–1836 (2003).



3. Pizette, S. & Niswander, L. BMPs negatively regulate structure and function of the limb apical ectodermal ridge. *Development* **126**, 883–894 (1999).
4. Scherz, P. J., Harfe, B. D., McMahon, A. P. & Tabin, C. J. The limb bud Shh-Fgf feedback loop is terminated by expansion of former ZPA cells. *Science* **305**, 396–399 (2004).
5. Aegerter-Wilmsen, T., Aegerter, C. M., Hafen, E. & Basler, K. Model for the regulation of size in the wing imaginal disc of *Drosophila*. *Mech. Dev.* **124**, 318–326 (2007).
6. Garcia-Bellido, A. C. & Garcia-Bellido, A. Cell proliferation in the attainment of constant sizes and shapes: the Entelechia model. *Int. J. Dev. Biol.* **42**, 353–362 (1998).
7. Hufnagel, L., Teleman, A. A., Rouault, H., Cohen, S. M. & Shraiman, B. I. On the mechanism of wing size determination in fly development. *Proc. Natl Acad. Sci. USA* **104**, 3835–3840 (2007).
8. Day, S. J. & Lawrence, P. A. Measuring dimensions: the regulation of size and shape. *Development* **127**, 2977–2987 (2000).
9. Sun, X. *et al.* Conditional inactivation of *Fgf4* reveals complexity of signalling during limb bud development. *Nature Genet.* **25**, 83–86 (2000).
10. Zuniga, A., Haramis, A. P., McMahon, A. P. & Zeller, R. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* **401**, 598–602 (1999).
11. Panman, L. *et al.* Differential regulation of gene expression in the digit forming area of the mouse limb bud by SHH and gremlin 1/FGF-mediated epithelial-mesenchymal signalling. *Development* **133**, 3419–3428 (2006).
12. Michos, O. *et al.* Gremlin-mediated BMP antagonism induces the epithelial-mesenchymal feedback signaling controlling metanephric kidney and limb organogenesis. *Development* **131**, 3401–3410 (2004).
13. Khokha, M. K., Hsu, D., Brunet, L. J., Dionne, M. S. & Harland, R. M. Gremlin is the BMP antagonist required for maintenance of Shh and Fgf signals during limb patterning. *Nature Genet.* **34**, 303–307 (2003).
14. Minowada, G. *et al.* Vertebrate Sprouty genes are induced by FGF signaling and can cause chondrodysplasia when overexpressed. *Development* **126**, 4465–4475 (1999).
15. Lewandoski, M., Sun, X. & Martin, G. R. Fgf8 signalling from the AER is essential for normal limb development. *Nature Genet.* **26**, 460–463 (2000).
16. Moon, A. M. & Capecchi, M. R. Fgf8 is required for outgrowth and patterning of the limbs. *Nature Genet.* **26**, 455–459 (2000).
17. Chiang, C. *et al.* Manifestation of the limb prepattern: limb development in the absence of sonic hedgehog function. *Dev. Biol.* **236**, 421–435 (2001).
18. Sun, X., Mariani, F. V. & Martin, G. R. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* **418**, 501–508 (2002).
19. Lu, P., Minowada, G. & Martin, G. R. Increasing Fgf4 expression in the mouse limb bud causes polysyndactyly and rescues the skeletal defects that result from loss of Fgf8 function. *Development* **133**, 33–42 (2006).
20. Xu, X., Qiao, W., Li, C. & Deng, C. X. Generation of Fgfr1 conditional knockout mice. *Genesis* **32**, 85–86 (2002).
21. Harfe, B. D. *et al.* Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
22. Eswarakumar, V. P. *et al.* The Ilc alternative of Fgfr2 is a positive regulator of bone formation. *Development* **129**, 3783–3793 (2002).
23. Merino, R. *et al.* The BMP antagonist Gremlin regulates outgrowth, chondrogenesis and programmed cell death in the developing limb. *Development* **126**, 5515–5522 (1999).
24. Nissim, S., Hasso, S. M., Fallon, J. F. & Tabin, C. J. Regulation of Gremlin expression in the posterior limb bud. *Dev. Biol.* **299**, 12–21 (2006).
25. Capdevila, J., Tsukui, T., Rodriguez Esteban, C., Zappavigna, V. & Izpisua Belmonte, J. C. Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. *Mol. Cell* **4**, 839–849 (1999).
26. Selever, J., Liu, W., Lu, M. F., Behringer, R. R. & Martin, J. F. Bmp4 in limb bud mesoderm regulates digit pattern by controlling AER development. *Dev. Biol.* **276**, 268–279 (2004).
27. Ovchinnikov, D. A. *et al.* BMP receptor type IA in limb bud mesenchyme regulates distal outgrowth and patterning. *Dev. Biol.* **295**, 103–115 (2006).
28. Bandyopadhyay, A. *et al.* Genetic analysis of the roles of BMP2, BMP4, and BMP7 in limb patterning and skeletogenesis. *PLoS Genet.* **2**, e216 (2006).
29. Pajni-Underwood, S., Wilson, C. P., Elder, C., Mishina, Y. & Lewandoski, M. BMP signals control limb bud interdigital programmed cell death by regulating FGF signaling. *Development* **134**, 2359–2368 (2007).
30. Mishina, Y., Hanks, M. C., Miura, S., Tallquist, M. D. & Behringer, R. R. Generation of Bmp4/Alk3 conditional knockout mice. *Genesis* **32**, 69–72 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to J. Fallon, G. Martin, R. Bacon, G. Boekhoff-Falk, B. Harfe, M. Lewandoski, D. Wellik and members of the Sun laboratory, in particular L. Abler, for discussions and reading the manuscript. We thank R. Behringer, C. Chiang, C. Deng, B. Harfe, R. Harland, P. Lonai, Y. Mishina and C. Tabin for mouse strains. We are grateful to A. Lashua, M. Zhao and J. Heinritz for technical assistance. J.M.V. was supported by the predoctoral training program in genetics (5T32GM07133) funded by the National Institutes of Health. This work was supported by a March of Dimes Basil O'Connor award 5-FY03-13 (to X.S.) and a National Institutes of Health grant RO1 HD045522 (to X.S.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to X.S. ([xsun@wisc.edu](mailto:xsun@wisc.edu)).

## METHODS

**Production of mutants.** Mutant embryos were genotyped using DNA from tail tissue and PCR primers in published reports of individual mutant alleles<sup>13,15,17,20–22,30</sup>. Using the matings described in Methods Summary, *Msx2cre;Fgf8<sup>del/flox</sup>;Shh<sup>-/-</sup>* embryos were generated at a predicted frequency of 1/8, and *Msx2cre;Fgf8<sup>flox/+</sup>;Shh<sup>+/-</sup>* embryos were used as controls. *Msx2cre;Fgf8<sup>del/flox</sup>;Gremlin<sup>-/-</sup>* embryos were generated at a predicted frequency of 1/16, and *Msx2cre;Fgf8<sup>flox/+</sup>;Gremlin<sup>+/-</sup>* embryos were used as controls. *Shh<sup>cre</sup>;Fgfr1<sup>co/co</sup>;Fgfr2<sup>c/c</sup>* embryos were generated at a predicted frequency of 1/8, and either *Fgfr1<sup>co/co</sup>;Fgfr2<sup>c/+</sup>* or *Fgfr1<sup>co/co</sup>;Fgfr2<sup>c/c</sup>* embryos were used as controls. In addition, *Prx1cre;r1* animals were generated by crossing *Prx1cre;Fgfr1<sup>co/+</sup>* males to *Fgfr1<sup>co/co</sup>* females to obtain *Prx1cre;Fgfr1<sup>co/co</sup>* animals at a frequency of 1/4. *Prx1cre;Fgfr1<sup>co/+</sup>* animals were used as controls. *Msx2cre;Fgf4;Fgf8* mutants and *Tcre;Fgfr1* mutants were generated as previously described<sup>18,31</sup>.

**Embryo isolation and phenotype analyses.** Embryos were dissected from time-mated mice, counting noon on the day the vaginal plug was found as E0.5. Wholemount *in situ* hybridization was performed as previously described<sup>32</sup>.

**Alcian blue skeletal preparations.** Embryos were fixed in 95% ethanol for two days, stained in 0.01% alcian blue, 0.015% alizarin red solution for three days, cleared in 1% KOH, photographed and stored in glycerol.

**Bead implantation experiments.** Heparin sulphate beads were soaked in either 1 mg ml<sup>-1</sup> or 0.1 mg ml<sup>-1</sup> Human FGF2 protein (PeproTech) at room temperature for 1 h, then implanted into the distal mesenchyme of stage 22 chick limb buds. Control beads were soaked in PBS. Embryos were harvested 12 h later and processed for RNA *in situ* hybridization. Contralateral limb buds were used as controls.

31. Verheyden, J. M., Lewandoski, M., Deng, C., Harfe, B. D. & Sun, X. Conditional inactivation of *Fgfr1* in mouse defines its role in limb bud establishment, outgrowth and digit patterning. *Development* **132**, 4235–4245 (2005).
32. Neubuser, A., Peters, H., Balling, R. & Martin, G. R. Antagonistic interactions between FGF and BMP signaling pathways: a mechanism for positioning the sites of tooth formation. *Cell* **90**, 247–255 (1997).

## LETTERS

# Amygdala intercalated neurons are required for expression of fear extinction

Ekaterina Likhtik<sup>1</sup>, Daniela Popa<sup>1</sup>, John Apergis-Schoute<sup>1</sup>, George A. Fidacaro Jr<sup>1</sup> & Denis Paré<sup>1</sup>

Congruent findings from studies of fear learning in animals and humans indicate that research on the circuits mediating fear constitutes our best hope of understanding human anxiety disorders<sup>1–4</sup>. In mammals, repeated presentations of a conditioned stimulus that was previously paired to a noxious stimulus leads to the gradual disappearance of conditioned fear responses. Although much evidence suggests that this extinction process depends on plastic events in the amygdala<sup>1–7</sup>, the underlying mechanisms remain unclear. Intercalated (ITC) amygdala neurons constitute probable mediators of extinction because they receive information about the conditioned stimulus from the basolateral amygdala (BLA)<sup>8,9</sup>, and contribute inhibitory projections to the central nucleus (CEA)<sup>10,11</sup>, the main output station of the amygdala for conditioned fear responses<sup>12</sup>. Thus, after extinction training, ITC cells could reduce the impact of conditioned-stimulus-related BLA inputs to the CEA by means of feed-forward inhibition. Here we test the hypothesis that ITC neurons mediate extinction by lesioning them with a toxin that selectively targets cells expressing  $\mu$ -opioid receptors ( $\mu$ ORs). Electron microscopic observations revealed that the incidence of  $\mu$ OR-immunoreactive synapses is much higher in ITC cell clusters than in the BLA or CEA and that  $\mu$ ORs typically have a post-synaptic location in ITC cells. In keeping with this, bilateral infusions of the  $\mu$ OR agonist dermorphin conjugated to the toxin saporin in the vicinity of ITC neurons caused a 34% reduction in the number of ITC cells but no significant cell loss in surrounding nuclei. Moreover, ITC lesions caused a marked deficit in the expression of extinction that correlated negatively with the number of surviving ITC neurons but not CEA cells. Because ITC cells exhibit an unusual pattern of receptor expression, these findings open new avenues for the treatment of anxiety disorders.

The hypothesis that ITC cells are involved in extinction<sup>13</sup> has not been tested yet because ITC cells occur as small, distributed cell clusters, making selective electrolytic or excitotoxic ITC lesions impossible. Here we circumvent this difficulty by exploiting the fact that ITC neurons express high levels of  $\mu$ ORs (Fig. 1a)<sup>14,15</sup>, allowing selective ITC lesions with a peptide–toxin conjugate that only targets  $\mu$ OR-expressing cells. Indeed, targeted toxins take advantage of receptor-mediated endocytosis to deliver cytotoxins to specific types of neurons<sup>16</sup>. Here, the peptide dermorphin, an agonist with a high affinity and selectivity for  $\mu$ ORs<sup>17</sup>, was conjugated to the ribosome-inactivating protein saporin (D-Sap).

For this lesion method to be effective,  $\mu$ ORs must be located postsynaptically in ITC cells, not in their afferents. Thus, we first used electron microscopy to determine their subcellular location. Using the electron microscope,  $\mu$ OR immunoreactivity was seen to be more concentrated in ITC cell clusters than in the BLA or the CEA (see Supplementary Fig. 1) and it was generally found postsynaptically (Fig. 1b, c and Supplementary Fig. 2). Indeed, the proportion of synapses that displayed postsynaptic  $\mu$ OR immunolabelling was

~3–6 times higher in the ITC cell clusters than in the CEA ( $P = 0.009$ ) or the BLA ( $P = 0.0001$ ;  $\chi^2$ -tests; Fig. 1c).

Having established that postsynaptic  $\mu$ OR expression is much higher in the ITC cell clusters than in neighbouring nuclei, we tested the feasibility of obtaining selective ITC lesions with the targeted toxin D-Sap and examined their impact on extinction. In brief, 58 rats were habituated to the training chamber (day 1) and were fear conditioned to a tone (day 2). Then, in a different context, they were trained on extinction (day 3). On day 4, they received bilateral infusions of either D-Sap (experimental group) or the same volume and concentration of a scrambled peptide conjugated to saporin (U-Sap, control group) aimed at ITC cells. On day 11, extinction recall was tested. The conditioned response we monitored was behavioural freezing, quantified by an observer blind to the rats' condition.

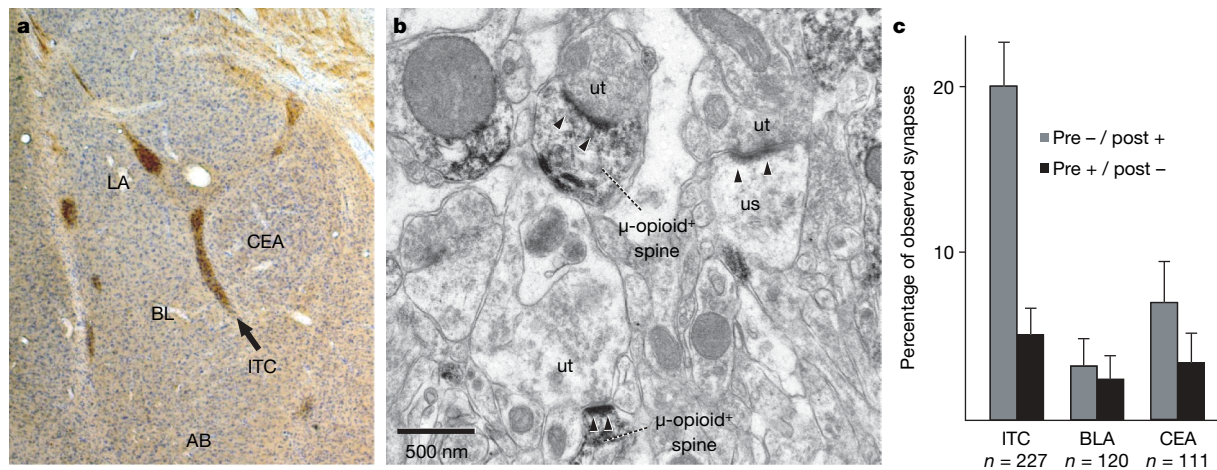
To maximize lesion specificity, we only considered rats in which the cannula tips were at the BLA–CEA border (that is, where CEA-inhibiting ITC cells are located), without knowledge of the behavioural data. In the control U-Sap and experimental D-Sap groups, 11 and 8 rats, respectively, met this criterion. Figure 2 shows coronal sections obtained from such control (Fig. 2a–c) and experimental (Fig. 2d–f) rats. Compared to control (U-Sap) cases, D-Sap infusions caused a marked but spatially circumscribed reduction in  $\mu$ OR staining restricted to the region adjacent to the infusion site (white arrows), where peri-CEA ITC clusters are normally found (black arrows). More distant ITC clusters such as those bordering the external capsule (arrowheads) or at the posterior pole of the amygdala (Fig. 2c, f) were not affected. Importantly, no difference in  $\mu$ OR labelling was seen in the CEA or BLA between experimental and control rats. Nevertheless, to control for possible effects of cell loss due to unintended D-Sap diffusion, we also analysed the behaviour of rats for which histological controls revealed that D-Sap infusions missed the BLA–CEA border but instead ended in the CEA ( $n = 8$ ; Fig. 2h) or the BLA ( $n = 7$ ; Fig. 2i).

To evaluate the neuronal loss caused by D-Sap infusions in ITC clusters, we performed unbiased stereological estimates of ITC and CEA neuronal numbers on sections counterstained with cresyl violet. The observer was blind to the rats' condition. Compared to U-Sap-treated rats, a significant reduction in the number of ITC cells was found in rats that received D-Sap infusions in ITC clusters (Fig. 2g; U-Sap,  $18,499 \pm 1,406$  (mean  $\pm$  s.e.m.),  $n = 8$ ; D-Sap,  $12,197 \pm 1,523$ ,  $n = 6$ ;  $-34\%$ ,  $t$ -test,  $P = 0.01$ ). In contrast, neuronal counts in CEA were nearly identical in the two groups (U-Sap,  $37,949 \pm 3,516$ ,  $n = 4$ ; D-Sap,  $40,471 \pm 2,328$ ,  $n = 5$ ;  $t$ -test,  $P = 0.55$ ).

Figure 3a shows the per cent of time that the rat spent freezing ( $y$  axis) during the different experimental phases in the three groups: the experimental group that received D-Sap infusions in the medial ITC clusters (red,  $n = 8$ ) and the two control groups that received either U-Sap infusions at the same site (open black circles,  $n = 11$ ) or D-Sap infusions centred on the BLA or CEA (filled black circles,  $n = 15$ ). Before D-Sap or U-Sap infusions, the three groups behaved similarly

<sup>1</sup>Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New Jersey, Newark, New Jersey 07102, USA.





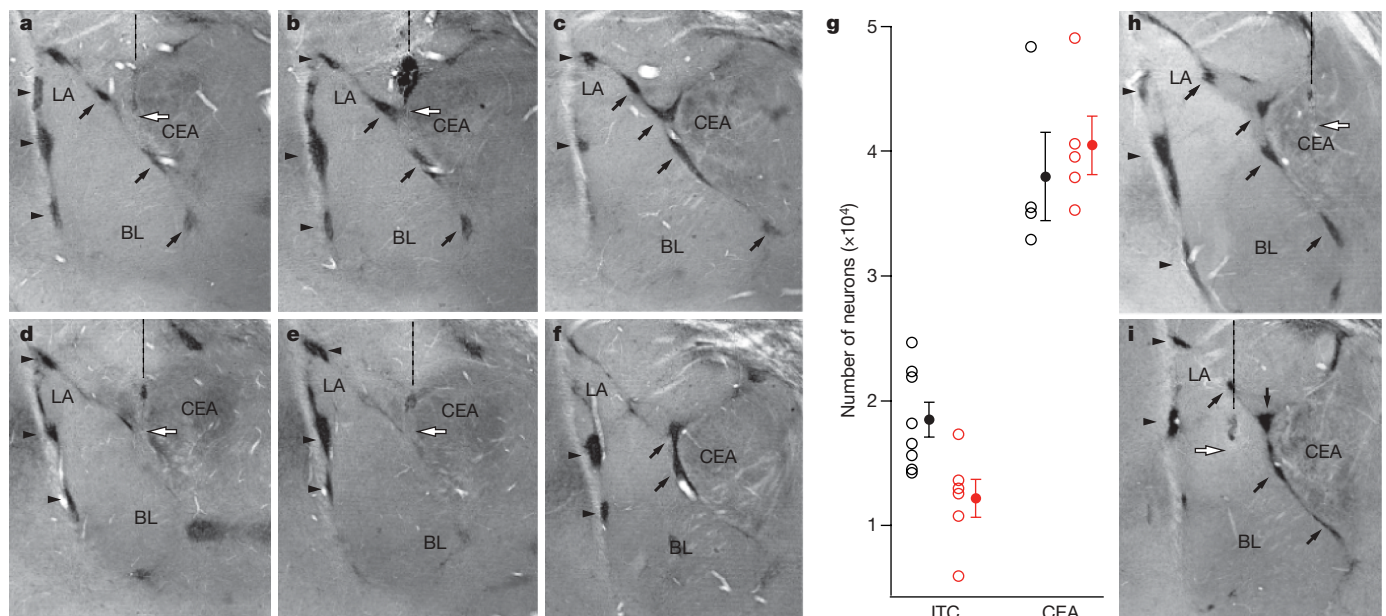
**Figure 1** |  $\mu$ OR immunoreactivity in the amygdala. **a**, Coronal section was processed to reveal  $\mu$ OR immunoreactivity (brown) and counterstained with cresyl violet (blue).  $\mu$ OR immunoreactivity is much higher in ITC cell clusters than in surrounding nuclei. AB, accessory basal nucleus; BL, basolateral nucleus; LA, lateral nucleus. **b**, Electron micrograph showing

examples of  $\mu$ OR<sup>+</sup> synapses in the ITC region. us, unlabelled spine; ut, unlabelled axon terminal. **c**, Proportion of synapses (mean  $\pm$  s.e.m.) at which  $\mu$ OR immunoreactivity was found in the postsynaptic (grey) or presynaptic (black) element in the ITC, BLA or CEA.

during habituation, fear conditioning and extinction training (Fig. 3a, days 1–3). However, after ITC lesions (day 4), their behaviour differed significantly during recall of extinction (Fig. 3a, day 11), with ITC-lesioned rats showing impaired expression of extinction (two-way repeated measures ANOVA (analysis of variance),  $F_{2,270} = 11.399$ ,  $P = 0.0002$ ). Post-hoc *t*-tests with stepwise Bonferroni correction of the significance level ( $P < 0.05$ ) comparing the D-Sap ITC group to the average of the two control groups in blocks of two trials revealed that rats with ITC lesions had significantly higher freezing levels during the first four conditioned stimuli (CS). Moreover, a strong inverse correlation was found between freezing levels during extinction testing and the number of surviving ITC cells (Fig. 3b, filled circles,  $r = -0.67$ ,  $P < 0.01$ ,  $n = 14$ ). In

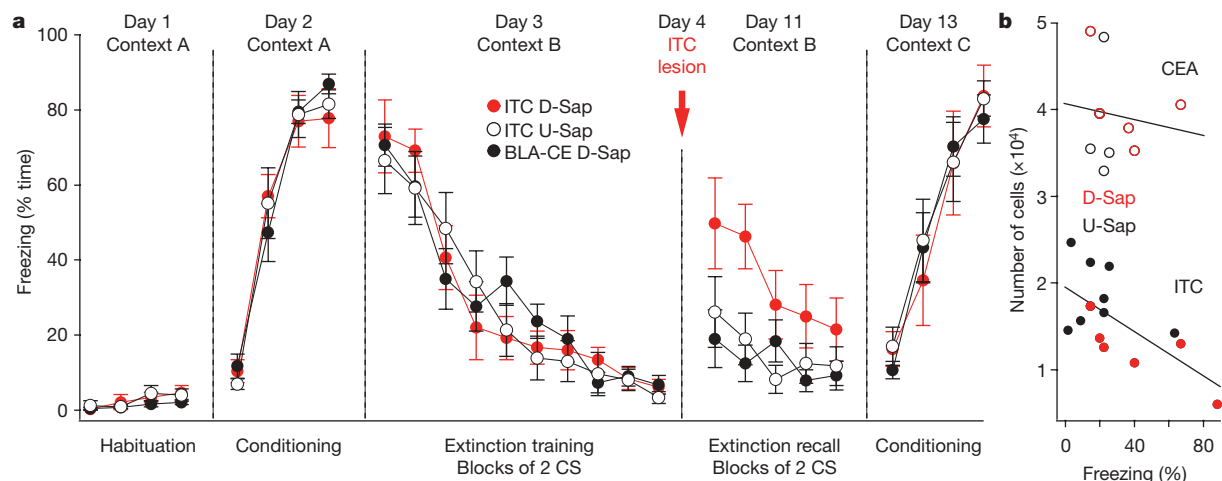
contrast, no such relationship was seen with CEA cell counts (Fig. 3b, open circles;  $r = -0.13$ ,  $P > 0.05$ ,  $n = 9$ ).

Inter-group differences were not attributable to non-specific increases in anxiety levels secondary to ITC lesions because exploratory behaviour in a new open field was indistinguishable between the three groups (ANOVAs,  $F_{2,25}$ : per cent time in centre  $F = 1.28$ ,  $P = 0.29$ ; distance travelled  $F = 1.11$ ,  $P = 0.34$ ). Moreover, the three groups acquired conditioned fear responses to a different CS at the same rate (Fig. 3, day 13; ANOVA,  $F_{2,93} = 0.007$ ,  $P = 0.9931$ ). Finally, the extinction deficit seen in the experimental group did not result from a non-specific CEA disinhibition, because freezing during the last CS of the first (day 2) and second (day 13) fear-conditioning sessions was identical (paired *t*-test,  $P = 0.61$ ).



**Figure 2** | D-Sap infusions at BLA–CEA border cause a spatially circumscribed loss of  $\mu$ OR immunoreactivity. **a–f**, Coronal sections obtained from rats that received either U-Sap (control, **a–c**) or D-Sap (experimental, **d–f**) infusions at the BLA–CEA border. The left and middle sections were obtained near the infusion site (white arrow); the right section was obtained near the caudal pole of the amygdala. Arrowheads and black

arrows point to lateral and medial ITC clusters, respectively. Dashed lines indicate the trajectory of the infusion cannulas. **g**, Unbiased stereological estimates of cell numbers (mean  $\pm$  s.e.m.) in the medial ITC clusters and CEA in control (black) versus experimental (D-Sap, red) animals. **h–i**, D-Sap infusion sites (white arrows) in the CEA (**h**) and the BLA (**i**).



**Figure 3 | D-Sap-induced ITC lesions cause an extinction deficit.** **a**, Per cent time freezing (y axis; mean  $\pm$  s.e.m.) in ITC-lesioned rats (red;  $n = 8$ ) compared to control rats that received U-Sap infusions at the same site (open black circles;  $n = 11$ ) or D-Sap infusions in the BLA or CEA (filled black circles;  $n = 15$ ). Data obtained in the D-Sap BLA and CEA animals were

pooled because the behaviour of these two rat subsets was statistically undistinguishable (ANOVA  $F_{1,108} = 0.847$ ,  $P = 0.38$ ). **b**, The relationship between the number of ITC (bottom) or CEA (top) neurons (y axis) and per cent time freezing (x axis) during the extinction testing session (average of the first three trials).

Although ITC lesions produced a deficit in extinction expression, the difference between the three groups progressively decreased with repeated CS presentations. Several factors probably contributed to the development of extinction after ITC lesions. First, much evidence indicates that the gradual fear reduction seen within an extinction training session depends on different mechanisms to those underlying between-session extinction. Indeed, some lesions<sup>3</sup> and pharmacological treatments<sup>6</sup> leave within-session extinction intact or marginally reduced, yet severely reduce between-session extinction. This implies that extinction engages at least two parallel processes: a first process that develops rapidly but does not last long, responsible for within-session extinction, and a second process that lasts longer and underlies between-session extinction. Although the mechanisms underlying within-session extinction remain unclear, a probable contributing factor is the progressive reduction in BLA unit responses as the CS is repeated. In addition, a participation of ITC cells to within-session extinction cannot be ruled out. Indeed, because our ITC lesions were incomplete, surviving ITC neurons might have contributed to within-session extinction. However, because the deficit caused by ITC cells was most pronounced at the onset of the testing session, ITC cells are probably critical to between-session extinction, as described below.

Extinction is known to depend on the reinforcement of an active  $\gamma$ -aminobutyric acid (GABA)ergic process<sup>18</sup> and on *N*-methyl-D-aspartic acid (NMDA)-dependent synaptic plasticity in the amygdala<sup>5–7,19</sup>. However, extinction training does not erase the initial fear memory because many BLA neurons maintain their increased CS responsiveness after extinction training<sup>20,21</sup>. Importantly, extinction training does not interfere with conditioned fear responses to a different CS or with the subsequent acquisition of conditioned fear responses to a different CS<sup>1</sup>.

We propose that ITC neurons can account for these properties of extinction. Indeed, ITC neurons receive CS information from the BLA<sup>8,9</sup> and send GABAergic projections to the CEA<sup>10,11</sup>. Thus, they are in a perfect position to regulate the flow of CS information from the BLA to the CEA<sup>9</sup>. Importantly, ITC neurons receive a dense projection from the infralimbic cortex<sup>22</sup>, the stimulation of which accelerates extinction<sup>23</sup> and inhibits CEA neurons<sup>24</sup>. We hypothesize that the CS-specificity of extinction derives from the ability of BLA synapses onto ITC neurons to express activity- and NMDA-dependent long-term potentiation<sup>13,25</sup>. During extinction training, convergence of CS-related infralimbic and BLA inputs onto ITC cells would facilitate induction of NMDA-dependent long-term potentiation,

but only at those BLA to ITC synapses recruited by the CS. This CS-specific potentiation of BLA inputs onto ITC cells would enhance the depolarization produced by the CS in ITC cells. Consequently, the GABAergic output of ITC cells onto CEA neurons would be increased, ultimately leading to a CS-specific reduction of conditioned fear responses. In light of recent findings showing increased bursting and tone responsiveness of infralimbic neurons after extinction training<sup>23,26</sup>, it is possible that the infralimbic cortex facilitates extinction-related plasticity during a consolidation phase.

Although the details of this model remain to be tested, the finding that ITC lesions produce a deficit in the expression of extinction that correlates negatively with the number of surviving ITC cells suggests that ITC cells are critically involved in extinction. The significance of this conclusion derives from earlier results suggesting that some human anxiety disorders reflect an extinction deficit<sup>3,4,27</sup> and functional imaging evidence that the medial prefrontal cortex and amygdala are hypo- and hyper-active, respectively, in such disorders<sup>28,29</sup>. As a result, it might be possible to compensate for these abnormalities and to facilitate extinction with pharmacological interventions that enhance the excitability of ITC cells by taking advantage of their unusual pattern of receptor expression<sup>14,15,30</sup>.

## METHODS SUMMARY

Procedures were approved by the Institutional Animal Care and Use Committee of Rutgers University, in compliance with the Guide for the Care and Use of Laboratory Animals (DHHS). For light and electron microscopic observations of  $\mu$ OR distribution, Sprague-Dawley rats were anaesthetized, perfused-fixed, their brains sectioned, and the sections processed as described in the Supplementary Information. Some sections were counterstained with cresyl violet and used for estimation of neuron numbers with the optical fractionator method. For electron microscopic observations, regions of interest were observed in four blocks obtained from two animals. For the behavioural study, 58 rats were subjected to a fear-conditioning paradigm as described in the Supplementary Information. One day after extinction training, the rats were anaesthetized, mounted in a stereotaxic apparatus, their skull exposed, a craniotomy performed, and drugs infused in the ITC region with a micro-syringe. Data are expressed as mean  $\pm$  s.e.m.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 27 February; accepted 10 June 2008.

Published online 9 July 2008.

- Myers, K. M. & Davis, M. Mechanisms of fear extinction. *Mol. Psychiatry* 12, 120–150 (2007).

2. Phelps, E. A. & LeDoux, J. E. Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* **48**, 175–187 (2005).
3. Quirk, G. J. & Mueller, D. Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* **33**, 56–72 (2008).
4. Ressler, K. J. & Mayberg, H. S. Targeting abnormal neural circuits in mood and anxiety disorders: from the laboratory to the clinic. *Nature Neurosci.* **10**, 1116–1124 (2007).
5. Falls, W. A., Miserendino, M. J. D. & Davis, M. Extinction of fear-potentiated startle: blockade by infusion of an NMDA antagonist into the amygdala. *J. Neurosci.* **12**, 854–863 (1992).
6. Sotres-Bayon, F., Bush, D. E. & LeDoux, J. E. Acquisition of fear extinction requires activation of NR2B-containing NMDA receptors in the lateral amygdala. *Neuropsychopharmacology* **32**, 1929–1940 (2007).
7. Walker, D. L., Ressler, K. J., Lu, K. T. & Davis, M. Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of D-Cycloserine as assessed with fear-potentiated startle in rats. *J. Neurosci.* **22**, 2343–2351 (2002).
8. Marowsky, A., Yanagawa, Y., Obata, K. & Vogt, K. E. A specialized subclass of interneurons mediates dopaminergic facilitation of amygdala function. *Neuron* **48**, 1025–1037 (2005).
9. Royer, S., Martina, M. & Paré, D. An inhibitory interface gates impulse traffic between the input and output stations of the amygdala. *J. Neurosci.* **19**, 10575–10583 (1999).
10. Paré, D. & Smith, Y. Distribution of GABA immunoreactivity in the amygdaloid complex of the cat. *Neuroscience* **57**, 1061–1076 (1993).
11. Paré, D. & Smith, Y. The intercalated cell masses project to the central and medial nuclei of the amygdala in cats. *Neuroscience* **57**, 1077–1090 (1993).
12. Davis, M. in *The Amygdala: A Functional Analysis* (ed. Aggleton, J. P.) 213–287 (Oxford Univ. Press, 2000).
13. Royer, S. & Paré, D. Bidirectional synaptic plasticity in intercalated amygdala neurons and the extinction of conditioned fear responses. *Neuroscience* **115**, 455–462 (2002).
14. Jacobsen, K. X., Hoistad, M., Staines, W. A. & Fuxe, K. The distribution of dopamine D1 receptor and mu-opioid receptor 1 receptor immunoreactivities in the amygdala and interstitial nucleus of the posterior limb of the anterior commissure: relationships to tyrosine hydroxylase and opioid peptide terminal systems. *Neuroscience* **141**, 2007–2018 (2006).
15. Poulin, J. F., Chevalier, B., Laforest, S. & Drolet, G. Enkephalinergic afferents of the centromedial amygdala in the rat. *J. Comp. Neurol.* **496**, 859–876 (2006).
16. Wiley, R. G. & Lappi, D. A. Targeted toxins in pain. *Adv. Drug Deliv. Rev.* **55**, 1043–1054 (2003).
17. Gaudriault, G., Nouel, D., Dal Farra, C., Beaudet, A. & Vincent, J. P. Receptor-induced internalization of selective peptidic mu and delta opioid ligands. *J. Biol. Chem.* **272**, 2880–2888 (1997).
18. Harris, J. A. & Westbrook, R. F. Evidence that GABA transmission mediates context-specific extinction of learned fear. *Psychopharmacology (Berl.)* **140**, 105–115 (1998).
19. Lee, H. & Kim, J. J. Amygdalar NMDA receptors are critical for new fear learning in previously fear-conditioned rats. *J. Neurosci.* **18**, 8444–8454 (1998).
20. Hobin, J. A., Goosens, K. A. & Maren, S. Context-dependent neuronal activity in the lateral amygdala represents fear memories after extinction. *J. Neurosci.* **23**, 8410–8416 (2003).
21. Repa, J. C. et al. Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nature Neurosci.* **4**, 724–731 (2001).
22. McDonald, A. J., Mascagni, F. & Guo, L. Projections of the medial and lateral prefrontal cortices to the amygdala: A *Phaseolus vulgaris* leucoagglutinin study in the rat. *Neuroscience* **71**, 55–75 (1996).
23. Milad, M. R. & Quirk, G. J. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* **420**, 70–74 (2002).
24. Quirk, G. J., Likhtik, E., Pelletier, J. G. & Paré, D. Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. *J. Neurosci.* **23**, 8800–8807 (2003).
25. Royer, S. & Paré, D. Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* **422**, 518–522 (2003).
26. Burgos-Robles, A., Vidal-Gonzalez, I., Santini, E. & Quirk, G. J. Consolidation of fear extinction requires NMDA receptor-dependent bursting in the ventromedial prefrontal cortex. *Neuron* **53**, 871–880 (2007).
27. Milad, M. R. et al. Presence and acquired origin of reduced recall for fear extinction in PTSD: Results of a twin study. *J. Psychiatr. Res.* **42**, 515–520 (2008).
28. Bremner, J. D., Elzinga, B., Schmahl, C. & Vermetten, E. Structural and functional plasticity of the human brain in posttraumatic stress disorder. *Prog. Brain Res.* **167**, 171–186 (2008).
29. Shin, L. M., Rauch, S. L. & Pitman, R. K. Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Ann. NY Acad. Sci.* **1071**, 67–79 (2006).
30. Fuxe, K. et al. The dopamine D1 receptor-rich main and paracapsular intercalated nerve cell groups of the rat amygdala: relationship to the dopamine innervation. *Neuroscience* **119**, 733–746 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by RO1 grant MH-073610 to D.P. and NRSA fellowship F31 MH76415 to E.L. from NIMH.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.P. ([pare@andromeda.rutgers.edu](mailto:pare@andromeda.rutgers.edu)).



## METHODS

**Histology.** To reveal  $\mu$ OR immunoreactivity, Sprague-Dawley rats were given an overdose of pentobarbital ( $100 \text{ mg kg}^{-1}$ , intraperitoneally) and perfused through the heart with 250 ml 0.9% saline ( $4^\circ\text{C}$ ), followed by 500 ml of paraformaldehyde (4%) and picric acid (0.2%) in PBS (0.1 M, pH 7.4,  $4^\circ\text{C}$ ). The brain was then extracted from the skull and a block containing the amygdala was prepared. The block was then glued to the stage of a vibrating microtome with a cyanoacrylate adhesive and cut in  $60\text{-}\mu\text{m}$  sections with a vibrating microtome. The sections were stored in PBS and then incubated in 1% sodium borohydride (in PBS) for 30 min, washed in PBS, pre-incubated in a blocking solution (10% goat serum, 1% BSA, Triton-X100, 0.3% for light microscopy, 0.04% for electron microscopy), and incubated overnight (around 16 h) in the primary antibody solution containing  $\mu$ OR antibody from ImmunoStar (1:4,000), 1% normal goat serum, 1% BSA, and Triton-X100 (0.3% and 0.04% for light versus electron microscopy) in PBS. Then, sections were incubated in the secondary antibody solution (Jackson, 1:200), followed by the avidin–biotin–complex (Vector). The immunoreactivity was then revealed with diaminobenzidine (Sigma) in tris(hydroxymethyl)-aminomethane (pH 7.6) buffer. A one-in-four series of sections was then counterstained with cresyl violet. Sections for electron microscopy were rinsed in phosphate buffer, post-fixed in 2% osmium tetroxide for one hour, rinsed in phosphate buffer, dehydrated in a graded series of alcohol and propylene oxide, embedded in resin, and baked for 48 h at  $60^\circ\text{C}$ . Blocks containing the regions of interest were sectioned at  $60\text{--}70 \text{ nm}$  using an ultramicrotome, collected onto copper mesh grids, and counterstained with 5% uranyl acetate and Reynold's lead citrate.  $\mu\text{OR}^+$  structures were easy to differentiate from unlabelled elements because of the electron-dense amorphous diaminobenzidine reaction product associated with them (Fig. 1b and Supplementary Figs 1–3).

**Fear conditioning and open field test.** We used Coulbourn conditioning chambers ( $25 \times 29 \times 28 \text{ cm}$ , with aluminium and Plexiglas walls), the appearance of which was modified by introducing salient sensory clues (see below). The conditioning chambers were placed in sound-attenuating boxes that contained a ventilation fan, and were illuminated by a single house light. On day 1, rats were habituated to the training chamber (context A) for 20 min. On day 2, they were presented with four-tone CS (4 kHz, 30 s), each co-terminating with a foot shock (unconditioned stimulus, US, 0.4 mA, 1 s). On day 3, rats were placed in a different conditioning box (context B) and presented with 20 CS (with no US). On day 11, their recall of extinction was probed with ten presentations of the CS alone in context B. On day 13, the rats were fear-conditioned to a different tone (1 kHz) in a different training context (context C). After testing in a novel open field, rats were perfused-fixed as described above for subsequent histology. Rats were subjected to this protocol in pairs, one from each group. The conditioned response we

monitored was behavioural freezing, quantified off-line with a stopwatch by an observer blind to the rats' identity. For fear conditioning (context A), we used the chamber described above. For extinction training (context B), we introduced a black plexiglass floor washed with peppermint soap in the chamber. For the second fear conditioning (context C), the chamber had striped plexiglass walls.

For the open field test, the rats were placed one by one in a new large plexiglass enclosure ( $120 \times 120 \text{ cm}$ ) with walls all around ( $35 \text{ cm}$  high). Their spontaneous exploratory behaviour was monitored with a video camera for 5 min. Attention was paid to the amount of time the rats spent along the walls versus in the centre of the open field as well as to the travelled distance and speed of travel.

**ITC lesions.** To be included in the lesion study, rats had to display  $\leq 15\%$  freezing at the end of the extinction training session. A total of 58 rats met this criterion. On day 4, rats were anaesthetized with isoflurane and were administered atropine methyl nitrate to reduce secretions and to aid breathing. The rats were then mounted in a stereotaxic apparatus. Betadine was applied to the scalp, and a local anaesthetic (bupivacaine, subcutaneously) was injected to the region to be incised. Ten minutes later, the skull was exposed, a craniotomy performed, and a micro-syringe (25-gauge) stereotactically directed to ITC clusters. Then, rats received bilateral infusions ( $0.01 \mu\text{l min}^{-1}$ ; total of  $0.25 \mu\text{l}$  per hemisphere) of either D-Sap (experimental;  $3 \text{ pmol } \mu\text{l}^{-1}$  in sterile saline) or the same volume and concentration of a scrambled peptide conjugated to saporin (U-Sap; control; Advanced Targeting systems). Ten minutes elapsed between the end of the infusion and the removal of the micro-syringe to minimize diffusion along the needle tract. The wound was then sutured and a local antibiotic applied.

**Cell counts.** In a 1-in-4 series of sections stained to reveal  $\mu\text{OR}$  immunoreactivity and counterstained with cresyl violet, the regions of interest were outlined using a microscope equipped with a motorized stage and StereoInvestigator software (MicroBrightField). Contour areas were measured with StereoInvestigator software, summed, and multiplied by the section thickness ( $60 \mu\text{m}$ ) and by the section sampling factor (4) to yield volume estimates. The optical fractionator method was used to estimate neuron numbers. In brief, neurons were counted in a 1-in-4 series of sections starting from a random position near the rostral pole of the amygdala. We counted neurons throughout the sections except for guard zones at the upper and lower surfaces ( $3 \mu\text{m}$ ). The optical dissector height was  $10.7 \mu\text{m}$ . To estimate the number of stained amygdala neurons, the regions of interest were sampled randomly and systematically (ITC counting frame,  $25 \times 25 \mu\text{m}$ ; grid size,  $45 \times 43 \mu\text{m}$ ; CEA, counting frame,  $35 \times 35 \mu\text{m}$ ; grid size,  $115 \times 115 \mu\text{m}$ ). Section thickness was re-measured each counting frame. For ITC counts, the Gundersen coefficients of error were  $0.5 \pm 0.01$  and  $0.04 \pm 0.01$  for the D-Sap and U-Sap groups, respectively. For the CEA counts, they were  $0.04 \pm 0.01$  for both groups.

## LETTERS

# Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors

Jeong Beom Kim<sup>1\*</sup>, Holm Zaehres<sup>1\*</sup>, Guangming Wu<sup>1</sup>, Luca Gentile<sup>1</sup>, Kinarm Ko<sup>1</sup>, Vittorio Sebastiano<sup>1</sup>, Marcos J. Araúzo-Bravo<sup>1</sup>, David Ruau<sup>2</sup>, Dong Wook Han<sup>1</sup>, Martin Zenke<sup>2</sup> & Hans R. Schöler<sup>1</sup>

Reprogramming of somatic cells is a valuable tool to understand the mechanisms of regaining pluripotency and further opens up the possibility of generating patient-specific pluripotent stem cells. Reprogramming of mouse and human somatic cells into pluripotent stem cells, designated as induced pluripotent stem (iPS) cells, has been possible with the expression of the transcription factor quartet Oct4 (also known as Pou5f1), Sox2, c-Myc and Klf4 (refs 1–11). Considering that ectopic expression of c-Myc causes tumorigenicity in offspring<sup>2</sup> and that retroviruses themselves can cause insertional mutagenesis, the generation of iPS cells with a minimal number of factors may hasten the clinical application of this approach. Here we show that adult mouse neural stem cells express higher endogenous levels of Sox2 and c-Myc than embryonic stem cells, and that exogenous Oct4 together with either Klf4 or c-Myc is sufficient to generate iPS cells from neural stem cells. These two-factor iPS cells are similar to embryonic stem cells at the molecular level, contribute to development of the germ line, and form chimaeras. We propose that, in inducing pluripotency, the number of reprogramming factors can be reduced when using somatic cells that endogenously express appropriate levels of complementing factors.

Mouse and human somatic cells can be reprogrammed into iPS cells by the expression of a defined set of factors (*Oct4*, *Sox2*, *c-Myc* and *Klf4*, as well as *Nanog* and human *LIN28*)<sup>1–11</sup>. It is possible to generate iPS cells from mouse and human fibroblasts in the absence of *c-Myc* retrovirus<sup>6,7</sup>, and therefore it was suggested that endogenous expression of c-Myc could have a role in the reprogramming process. Neural stem cells (NSCs) endogenously express Sox2, which may function in maintaining the stemness and multipotency of NSCs<sup>12,13</sup>, and Sox2 was suggested in maintaining cellular pluripotency by regulating the expression of Oct4 (ref. 14). NSCs were established from adult Oct4–GFP (OG2)/*ROSA26* heterozygous transgenic mouse brains<sup>15–17</sup>, expressing green fluorescent protein (GFP) under the control of the *Oct4* promoter (Oct4–GFP) and the *Escherichia coli lacZ* transgene from the constitutive *ROSA26* locus (*ROSA26 lacZ*). The established NSCs had the capacity for self-renewal and multipotency (Supplementary Fig. 1).

Compared to embryonic stem cells (ESCs), expression of Sox2 was approximately twofold higher in NSCs. *c-Myc* and *Klf4* were also expressed at levels about tenfold higher and eightfold lower in NSCs than in ESCs, respectively (Fig. 1a). Western blot analyses showed that the relationship between protein and RNA levels in NSCs corresponded to that in ESCs for Oct4, Sox2 and Klf4; the c-Myc protein level was comparable in NSCs and ESCs (Fig. 1b).

In this study, we attempted to reprogram NSCs into iPS cells by introducing different combinations of the four factors *Oct4*, *Sox2*,

*c-Myc* and *Klf4* (Supplementary Table 1) using the retroviral MX vector system. We assessed the ability of 15 different transcription factor combinations to induce Oct4–GFP-positive colony formation. Six combinations were able to induce the generation of iPS cells from NSCs, as judged by the formation of GFP<sup>+</sup> colonies and the establishment of iPS cell lines. We observed GFP<sup>+</sup> cells 4 days after transduction with the combination containing all four factors—that is, the control combination—and noted a gradual increase in the number of GFP<sup>+</sup> colonies during the first 2 weeks post-infection (Supplementary Fig. 2a). We established four-factor iPS cells from GFP<sup>+</sup> ESC-like colonies on day 14. These four-factor iPS cells stained positive for stage-specific embryonic antigen-1 (SSEA-1) and alkaline phosphatase (Supplementary Fig. 2b), showed messenger RNA expression patterns similar to those in ESCs (Fig. 2a, Supplementary Fig. 2c), and led to teratoma formation on injection into nude mice (Supplementary Fig. 2d). Our results demonstrate that NSCs can be reprogrammed into iPS cells by the four transcription factors: Oct4, Sox2, c-Myc and Klf4.

Three different combinations of three factors were also capable of generating iPS cells from NSCs: *Oct4*, *Klf4* and *c-Myc* (OKM); *Oct4*, *Klf4* and *Sox2* (OKS); and *Oct4*, *c-Myc* and *Sox2* (OMS; Supplementary Table 1). We did not observe GFP<sup>+</sup> colonies for the three-factor combinations that did not include Oct4. GFP<sup>+</sup> colonies were observed 1 week after transduction with the OKM combination (without *Sox2*). However, GFP<sup>+</sup> colony formation was observed only after 14–21 days with the OKS combination (without *c-Myc*), and was even more delayed with the OMS combination (without *Klf4*; Supplementary Table 1). Nonetheless, these OKM, OKS and OMS three-factor iPS cells had similar gene expression profiles to ESCs (Supplementary Fig. 3a), and all types of three-factor iPS cells differentiated into all three germ layers (Supplementary Fig. 3b). Taken together, these results indicate that three-factor iPS cells could be generated in the absence of *Sox2*, *Klf4* or *c-Myc* retroviruses in NSCs, which endogenously express these three factors.

We next assessed the ability of two-factor combinations to induce the generation of iPS cells from NSCs. Only two combinations were successful in reprogramming NSCs. We first observed GFP<sup>+</sup> colonies in NSC cultures infected with *Oct4* and *Klf4* (OK) and 1–2 weeks later in those infected with *Oct4* and *c-Myc* (OM; Supplementary Table 1). The two-factor OM iPS cells showed an ESC-like gene expression pattern and contributed to the three germ layers in teratomas (Supplementary Fig. 3a, b). Klf4 and c-Myc may exert similar functions. It is known that Klf4 has a role in the inactivation of the *p53* (also known as *Trp53*) tumour-suppressor gene, which leads to cell immortalization; Klf4 also works in conjunction with the RAS<sup>V12</sup> oncogenic signal transduction protein to stimulate cellular proliferation<sup>18,19</sup>.

<sup>1</sup>Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Röntgenstrasse 20, 48149 Münster, NRW, Germany. <sup>2</sup>Institute for Biomedical Engineering, Department of Cell Biology, RWTH Aachen University Medical School, Pauwelsstrasse 30, 52074 Aachen, NRW, Germany.

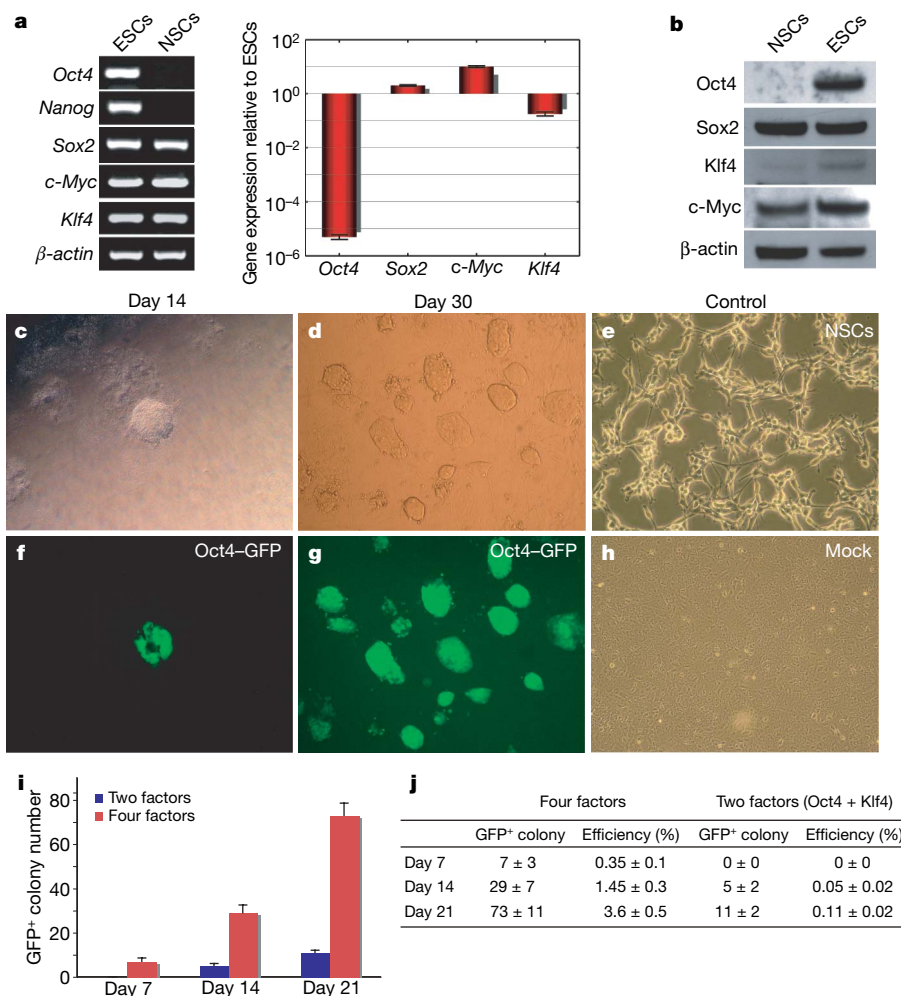
\*These authors contributed equally to this work.

Similarly, the immortalizing gene product *c-Myc*, in conjunction with mutant *RAS*, exhibits an oncogenic effect<sup>20</sup>. It has been reported that *c-Myc* increases telomerase activity in NSCs, a mechanism possibly responsible for the immortalization of NSCs<sup>21</sup>. Because *c-Myc* increases tumorigenicity in chimaera pups<sup>2</sup>, the recent studies demonstrating iPS cell generation without the *c-Myc* retrovirus<sup>6,7</sup> represent a significant finding. Importantly, our results of inducing iPS cells from NSCs with *Oct4* and *Klf4* and without *c-Myc* bring us even closer to the generation of iPS cells suitable for transplantation.

We compared two-factor OK iPS cells with four-factor iPS cells and ESCs. On day 14 post-infection, five GFP<sup>+</sup> colonies were dissociated and propagated under ESC culture conditions (Fig. 1c, f), yielding three (that is, 60%) two-factor OK iPS cell clones (B-2, D-7 and F-4) that were morphologically indistinguishable from ESCs (Fig. 1d, g). No colonies formed from NSCs infected with control virus (MX) (Fig. 1e, h). We estimated the reprogramming efficiencies from the number of Oct4–GFP<sup>+</sup> colonies and transduction rates with MX–GFP control virus on NSCs for the two-factor OK iPS and the four-factor iPS by time course (Fig. 1i, j). Thereby, we calculated a reprogramming efficiency of 3.6% for four-factor reprogramming of NSCs and 0.11% for the two-factor approach, which is comparable to reprogramming of fibroblasts with selection (<0.08%)<sup>1–3</sup> and

without selection (0.5%)<sup>5</sup> (Fig. 1j and Supplementary Table 2). Transduction with all four factors had a positive impact on the timing and number of GFP<sup>+</sup> colonies. Integration of the viral transgenes was confirmed by genotyping by polymerase chain reaction (PCR; Supplementary Fig. 4). The viral transgenes of all four factors were detected in four-factor iPS cells, whereas two-factor OK iPS cells only contained the *Oct4* and *Klf4* transgenes.

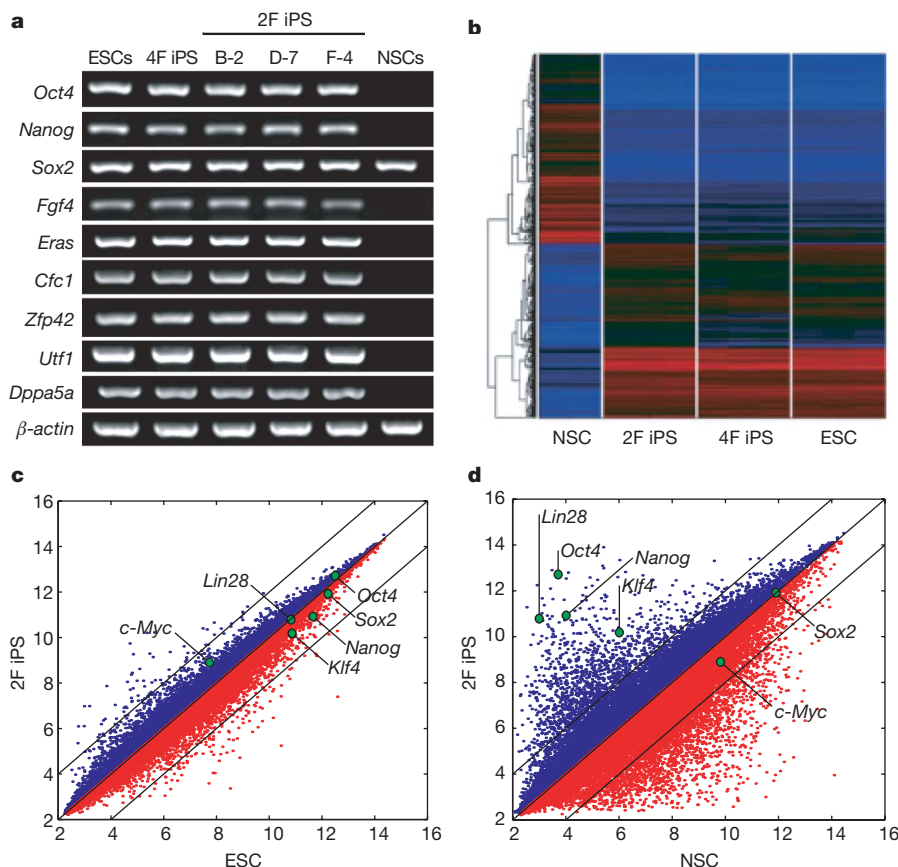
Two-factor OK iPS cells stained positive for SSEA-1 and alkaline phosphatase (Supplementary Fig. 5), and exhibited expression patterns of ESC marker genes similar to those in four-factor iPS cells and ESCs (Fig. 2a). Quantitative real-time PCR (qRT-PCR) results demonstrated that expression of endogenous *Oct4*, *Sox2*, *c-Myc* and *Klf4* in two-factor OK iPS cells was comparable to that in ESCs (Supplementary Fig. 6a), and showed the silencing of the viral transcripts in two-factor OK iPS cells with a 1,000-fold reduction after 30 days (Supplementary Fig. 6b). In addition, in four-factor iPS cells, the endogenous expression levels were similar to those in ESCs, and the expression of the transgenes was completely silenced (Supplementary Fig. 6c, d). Global gene expression of two-factor iPS also clustered close to ESCs and four-factor iPS (Fig. 2b and Supplementary Fig. 7). Scatter plots of DNA microarray analyses demonstrated a higher similarity between two-factor iPS cells and ESCs than between two-factor



**Figure 1 | Generation of two-factor Oct4/Klf4 (OK) iPS cells from adult NSCs of OG2/ROSA26 transgenic mice.** **a**, RT–PCR and qRT–PCR analyses of *Oct4*, *Nanog*, *Klf4*, *Sox2* and *c-Myc* in ESCs and NSCs ( $n=3$ ; error bars indicate s.d.).  $\beta$ -actin was used as a loading control. **b**, Western blot analyses of the four factors in ESCs and NSCs. Anti-actin antibody was used as a loading control. **c**, Morphology of two-factor OK iPS cell colony on day 14 post-infection. Shown is an ESC-like colony expressing Oct4–GFP (f). **d, g**, Morphology of established two-factor OK iPS cells (clone F-4) on

day 30 post-infection, grown on irradiated MEFs. Phase contrast (**d**) and Oct4–GFP (**g**) are shown. **e, h**, Morphology of NSCs and mock infection on day 30 post-infection. **i**, Generation of GFP-positive colonies at day 7, 14 and 21 after two-factor OK and four-factor infection ( $n=3$ ; error bars indicate s.d.). **j**, Reprogramming efficiency of generating two-factor and four-factor iPS cells ( $n=3$ ). Indicated are the total numbers of GFP<sup>+</sup> colonies per 50,000 plated NSCs on day 7, 14 and 21 after infection.





**Figure 2 | Gene expression profile of iPS cells.**

**a**, RT-PCR analysis of ESC marker gene expression in ESCs, four-factor (4F) iPS cells (clone A-2c), two-factor (2F) OK iPS cells (clones B-2, D-7 and F-4), and NSCs. Primers are specific for transcripts from the respective endogenous locus.  $\beta$ -actin was used as a loading control.

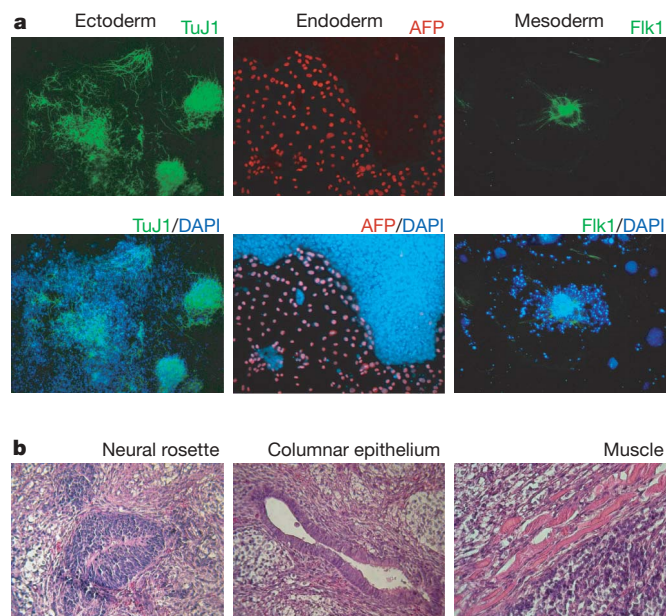
**b**, The heat map of the different expressed genes among the NSC, 2F (OK) iPS, 4F iPS and ESCs. The gene hierarchical cluster was performed with a cityblock distance and an average linkage. Red, high gene expression; blue, low gene expression.

**c, d**, Global gene expression patterns were compared between 2F iPS cells (clone F-4) and ESCs (**c**), and between 2F iPS cells (clone F-4) and NSCs with DNA microarrays (**d**). Black lines indicate twofold changes in gene expression levels between the paired cell types. Genes overexpressed in 2F iPS cells (clone F-4) compared with NSCs or ESCs are shown in blue; those underexpressed are shown in red. Positions of pluripotency genes *Oct4*, *Nanog*, *Sox2*, *c-Myc*, *Klf4* and *Lin28* in scatter plots are indicated. The gene expression levels were normalized with the RMA algorithm<sup>26</sup>.

iPS cells and NSCs (Fig. 2c, d). Thus, two-factor iPS cells (clone F-4) seemed to be very similar to mouse ESCs at the global transcription level.

The differentiation ability of two-factor OK iPS cells was confirmed by *in vitro* differentiation into embryoid bodies. These cells expressed the ectoderm (Tuj1), endoderm ( $\alpha$ -fetoprotein) and mesoderm (Flk1) markers (expressed by beating cells mimicking cardiomyocytes; Fig. 3a and Supplementary Video 1). Teratomas contained derivatives of all three germ layers (Fig. 3b) and expressed markers of the three germ layers (Supplementary Fig. 8). No teratoma had formed from donor cells (NSCs). These data demonstrate that two-factor OK iPS cells exhibit a pluripotent phenotype *in vitro* and *in vivo*.

To investigate their developmental potential, two-factor OK iPS cells were aggregated with 8-cell-stage embryos. iPS cells had contributed to the formation of the inner cell mass in developing blastocysts (Fig. 4a). After transferring aggregated blastocysts into pseudopregnant females, we obtained 16 live embryos on embryonic day (E)13.5, 2 of which showed germ cell contribution in the foetal gonads, judged from Oct4-GFP expression (Fig. 4b).  $\beta$ -galactosidase staining (visualizing the NSC donor cells that carry the *ROSA26 lacZ* gene) of embryonic tissue from whole embryos revealed that in the resulting chimaeras two-factor OK iPS cells contributed to the development of all three germ layers (Fig. 4c, e). The strictest test for developmental potency, tetraploid (4N) embryo aggregation ( $n = 122$ ), resulted in 2 dead (arrested) embryos at E13.5 (Fig. 4d). These data demonstrate that iPS cells can give rise to all of the tissues of a late-stage embryo. In diploid (2N) aggregation, PCR genotyping showed that 2 out of 13 chimaeras were positive for the Oct4-GFP allele of donor cells (Fig. 4f, g, top panel). To assess whether two-factor OK iPS cells can contribute to the germ line, chimaeras were mated with CD-1 females. Two out of 12 pups had a Oct4-GFP allele and 1 out of 12 mice had a *lacZ* allele, because the donor cells are derived from a heterogeneous mouse (*Oct4-GFP*<sup>+/−</sup> *ROSA26*<sup>+/−</sup>),



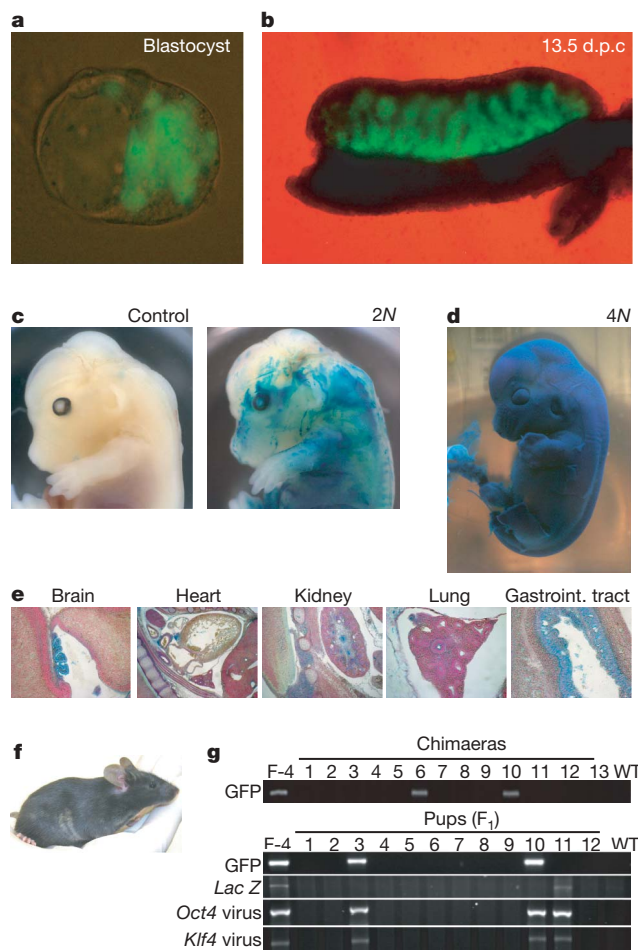
**Figure 3 | Two-factor Oct4/Klf4 (OK) iPS cells (clone F-4) are pluripotent and differentiate *in vitro* and *in vivo*.**

**a**, *In vitro* differentiation into all three germ layers. After embryoid body formation, aggregates were transferred onto gelatine-coated plates and allowed to differentiate for another 10 days. Cells were stained with anti-Tuj1, anti- $\alpha$ -fetoprotein (AFP) or anti-Flk1. Nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI).

**b**, Teratomas of F-4 iPS cells containing all three germ layers. F-4 iPS cells ( $1.5 \times 10^6$  cells) were subcutaneously inoculated into nude mice. After 4 weeks, teratomas were stained with haematoxylin and eosin dyes. Shown is a teratoma containing a neural rosette (ectoderm), columnar epithelium (endoderm) and muscle (mesoderm).

and they also had the *Oct4* and *Klf4* transgenes (Fig. 4g, bottom panel). No tumour formation was observed from adult chimaeras and F<sub>1</sub> mice by the age of 21 weeks and 7 weeks, respectively. This finding indicates that two-factor OK iPS cells can contribute the full-term development of chimaera, resulting in a next generation (F<sub>1</sub>) of viable pups, and thus suggests that the iPS cells have a similar developmental property to that of ESCs.

In summary, our study demonstrates that Oct4 together with either *Klf4* or c-Myc can reprogram NSCs into pluripotent stem cells. This is the first demonstration of an ectodermal cell type to be reprogrammed by defined factor overexpression. We propose that endogenous expression of Sox2 plus exogenous expression of two factors including Oct4 is sufficient to induce the generation of iPS cells from adult NSCs. However, it is not clear why exogenous expression of either *Klf4* or c-Myc is still required for reprogramming, because these genes are endogenously expressed in NSCs.



**Figure 4 | In vivo developmental potential of two-factor Oct4/Klf4 (OK) iPS cells (clone F-4).** **a**, The chimaeric embryos of F-4 iPS cells developed to blastocysts after 24 h of aggregation. Fluorescence optics show Oct4-GFP cells located in the inner cell mass of blastocysts. **b**, Germline contribution of F-4 iPS cells to mouse embryonic development as shown by the expression of Oct4-GFP. Embryos were analysed with a fluorescence microscope at E13.5. **c**, **d**, The 13.5 days post coitum (d.p.c.) chimaeric embryos (control, 2N and 4N) were stained with X-gal solution. **e**, Histological analysis of lacZ-stained 13.5 d.p.c. chimaeric embryo (2N). Gastroint. tract, gastrointestinal tract. **f**, Chimeric mouse (8-week-old) generated by F-4 iPS cells. The agouti coat colour originated from F-4 iPS cells. **g**, PCR genotyping of chimaeras derived from an F-4 iPS cell. PCR analyses were performed for Oct4-GFP (top panel). Germline transmission of F-4 iPS cells. Genotyping of offspring from chimaeric males mated with CD-1 females demonstrated the presence of Oct4-GFP and *lacZ* allele as well as *Oct4* and *Klf4* virus integrations (bottom panel). WT, wild type.

Human NSCs might be available from patient biopsy<sup>22</sup>, and it will be particularly interesting to assess whether they can be reprogrammed with the currently described two-factor method. The clinical applicability of cellular reprogramming is incumbent on the optimization of approaches that obviate problems caused by retroviral insertional mutagenesis. Chromatin-modifying agents, such as trichostatin A and 5-Aza-2'-deoxycytidine, induce the transient expression of several pluripotency-associated genes, including *Oct4* and *Klf4* (ref. 23). This indicates that one or more reprogramming factor(s) could be replaced by such chemical compounds and, in this regard, the two-factor transcription factor reprogramming induction model holds great potential.

## METHODS SUMMARY

NSCs were isolated from OG2/ROSA26 (Oct4-GFP) transgenic mice and transduced with 15 combinations of retroviral vectors expressing Oct4, Sox2, *Klf4* and c-Myc as described<sup>1,24</sup>. Oct4-GFP-positive colonies were mechanically isolated and cultured on irradiated MEFs in ESC medium without any further selection. iPS cells were characterized by qRT-PCR and gene expression profiling by microarray. *In vitro* differentiation by embryoid body formation, the teratoma assay on injection into nude mice and chimaera formation by aggregation with 8-cell-stage mouse embryos were used to measure the developmental potential of the iPS cells.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 20 March; accepted 8 May 2008.

Published online 29 June 2008.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
2. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
3. Wernig, M. *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
4. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
5. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotechnol.* **25**, 1177–1181 (2007).
6. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnol.* **26**, 101–106 (2008).
7. Wernig, M., Meissner, A., Cassady, J. P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cells* **2**, 11–12 (2008).
8. Aoi, T. *et al.* Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* doi:10.1126/science.1154884; published online 14 February 2008.
9. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
10. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
11. Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
12. Episkopou, V. SOX2 functions in adult neural stem cells. *Trends Neurosci.* **28**, 219–221 (2005).
13. Miyagi, S. *et al.* The Sox-2 regulatory regions display their activities in two distinct types of multipotent stem cells. *Mol. Cell. Biol.* **24**, 4207–4220 (2004).
14. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biol.* **9**, 625–635 (2007).
15. Ryan, A. K. & Rosenfeld, M. G. POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev.* **11**, 1207–1225 (1997).
16. Do, J. T. & Scholer, H. R. Nuclei of embryonic stem cells reprogram somatic cells. *Stem Cells* **22**, 941–949 (2004).
17. Pollard, S. M., Conti, L., Sun, Y., Goffredo, D. & Smith, A. Adherent neural stem (NS) cells from fetal and adult forebrain. *Cereb. Cortex* **16** (Suppl 1), i112–i120 (2006).
18. Sherr, C. J. & McCormick, F. The RB and p53 pathways in cancer. *Cancer Cell* **2**, 103–112 (2002).
19. Rowland, B. D., Bernards, R. & Peeper, D. S. The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene. *Nature Cell Biol.* **7**, 1074–1082 (2005).
20. Peeper, D. S. *et al.* A functional screen identifies hDR11 as an oncogene that rescues RAS-induced senescence. *Nature Cell Biol.* **4**, 148–153 (2002).
21. Miura, T. *et al.* Neural stem cells lose telomerase activity upon differentiating into astrocytes. *Cytotechnology* **36**, 137–144 (2001).

22. Nunes, M. C. *et al.* Identification and isolation of multipotential neural progenitor cells from the subcortical white matter of the adult human brain. *Nature Med.* **9**, 439–447 (2003).
23. Ruau, D. *et al.* Pluripotency associated genes are reactivated by chromatin modifying agents in neurosphere cells. *Stem Cells* **26**, 920–926 (2008).
24. Zaehres, H. & Daley, G. Q. Transgene expression and RNA interference in embryonic stem cells. *Methods Enzymol.* **420**, 49–64 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Müller-Keuker for critically reviewing the manuscript, M. Stehling for fluorescence-activated cell sorting analysis, C. Ortmeier for real-time PCR analysis, B. Schäfer for histology, C. Becker, B. Kratz and B. Denicke for probe processing and array hybridization, and T. Kitamura for the pMX retroviral vector. J.B.K. was supported by the Interdepartmental

Graduate-Programme for Experimental Life Sciences (iGEL) at the University of Münster. The microarray analyses were funded in part by a grant from the Deutsche Forschungsgemeinschaft DFG SPP1109.

**Author Contributions** J.B.K. (project design, generation and characterization of iPS cells and NSCs, and preparation of manuscript), H.Z. (project design, generation of iPS cells, and preparation of manuscript), G.W., L.G., V.S., K.K. and D.W.H. (characterization of iPS cells), M.J.A.-B., D.R., M.Z. (microarray and bioinformatics) and H.R.S. (project design, and preparation of manuscript).

**Author Information** The microarray data sets are available from the GEO (Gene Expression Omnibus) website under accession number GSE10806. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.R.S. ([schoeler@mpi-muenster.mpg.de](mailto:schoeler@mpi-muenster.mpg.de)).



## METHODS

**Mice.** The OG2 strain was crossed with the ROSA26 transgenic strain<sup>16,25</sup> over several generations to produce compound homozygous mice for the *neo/lacZ* and *Oct4-GFP* transgenes. To derive NSCs, homozygous OG2 × ROSA26 male mice were crossed with ICR mouse strain females to produce heterozygous pups. Brain tissue was collected from 5-day-old OG2 × ROSA26 heterozygous mice.

**Generation of induced pluripotent stem cells.** iPS cells and ESCs were grown on irradiated MEFs and in ESC medium (DMEM supplemented with 15% FBS, non-essential amino acids, L-glutamine, penicillin/streptomycin, β-mercaptoethanol, and 1,000 Units ml<sup>-1</sup> leukaemia inhibitory factor (LIF)). pMX-based retroviral vectors encoding the mouse complementary DNAs of *Oct4*, *Sox2*, *Klf4* and *c-Myc*<sup>1</sup> were separately cotransfected with packaging-defective helper plasmids into 293 cells using Eugene 6 transfection reagent (Roche). Forty-eight hours later, virus supernatants were collected as described previously<sup>24</sup>. NSCs derived from OG2/ROSA26 transgenic mice were seeded at a density of  $5 \times 10^4$  cells per 6-well plate and incubated with virus-containing supernatants for the four factors (1:1:1:1) or for *Oct4* and *Klf4* (1:1) supplemented with  $6 \mu\text{g ml}^{-1}$  protamine sulphate (Sigma) for 24 h. Transduction efficiencies were calculated with the pMX-GFP control virus. Cells were re-plated in fresh neural expansion medium. Two days after infection, the cells were further subcultured on irradiated MEFs in ESC medium containing LIF without any further selection. Oct4-GFP-positive colonies were mechanically isolated, and individual cells were dissociated and subsequently replated onto MEFs. The colonies were selected for expansion.

**qRT-PCR analysis.** Total RNA was extracted from cells using the MiniRNasey Kit (Qiagen GmbH; <http://www.qiagen.com>) according to the manufacturer's instructions. Complementary DNA synthesis was performed with the High Capacity cDNA Archive Kit (Applied Biosystems GmbH; <http://www.applied-biosystems.com>) following the manufacturer's instructions with a down-scaled reaction volume of 20 μl. Transcript levels were determined using the ABI PRISM Sequence Detection System 7900 (Applied Biosystems) and the ready-to-use 5'-nuclease Assays-on-Demand. For each real-time amplification, the template was equivalent to 5 ng of total RNA. Measurements were performed in triplicate; a reverse-transcription-negative blank of each sample and a no-template blank served as negative controls. Amplification curves and gene expression were normalized to the housekeeping gene *Hprt1*, used as an internal standard.

Oligonucleotides were designed by the Taqman Assay-on-Demand for the detection of the following genes (Taqman primer ID): *Oct4* (Mm00658129\_gH), *Sox2* (Mm00488369\_s1), *c-Myc* (Mm00487803\_m1), *Klf4* (Mm00516104\_m1) β-actin (Mm00607939\_s1) and *Hprt1* (Mm00446968\_m1). Oligonucleotides for the detection of *Nanog* and the viral sequences were custom-designed. Quantification was normalized on the endogenous *Hprt1* gene within the log-linear phase of the amplification curve obtained for each probe/primers set using the ΔΔCt method (ABI PRISM 7700 Sequence Detection System, user bulletin number 2).

**Primer sequences for viral-specific qRT-PCR.** pMXs-Oct4 PF: 5'-TGGTACGGGAAATCACAAGTTTG, PR: 5'-GTCATAGTTC CTGT TGGTGA AGTTCA, probe: 5'-6FAM-CTTCACCATGCCCCCTCA-MGB. pMXs-Sox2 PF: 5'-GTGTGGTGGTACGGGAAATCAC, PR: 5'-TTCAGCTCCGT CTC ATCATG, probe: 5'-6FAM-TGTACAAAAAGCAGGCTTGT-MGB. pMXs-Klf4 PF: 5'-GTGTGGTGGTACGGGAAATCA, PR: 5'-CGCGAACGTGG AGAA GGA, probe: 5'-6FAM-CTTCACCATGGCTGTCAG-MGB. pMXs-cMyc PF: 5'-TGGTACGGGAAATCACAAGTTTG, PR: 5'-GTCATAGTTC

CTGTTGGTGAAGTTCA, probe: 5'-6FAM-CTTCACCATGCCCCCTCA-MGB. Nanog PF: 5'-AACCAGTGGTTGAATACTAGCAATG, PR: 5'-CTGCAATGGAT GCTG GGATACT, probe: 5'-6FAM-TTCAGAA GGGCTCAGCAC-MGB.

**Microarray analysis.** The microarray study was carried out using Mouse Genome 430 2.0 GeneChip arrays (Affymetrix) essentially as described previously<sup>23</sup>. In brief, total RNA was extracted from cells with the RNeasy kit including DNase digestion (Qiagen). Biotin-labelled cRNA was obtained from 3 μg of total RNA with the GeneChip One-Cycle labelling kit (Affymetrix). Fifteen micrograms of cRNA were fragmented and hybridized to Affymetrix 430 2.0 GeneChip arrays at 45 °C for 16 h. DNA chips were washed, stained and scanned using an Affymetrix Fluidics device and a GCS3000 scanner, and the images obtained were analysed using the GCOS software. The experiment was performed in triplicates for the ESCs and iPS cells and in duplicates for the NSCs. Normalization was calculated with the RMA algorithm<sup>26</sup> implemented in BioConductor.

**In vitro differentiation of iPS cells.** Oct4-GFP cells were harvested by fluorescence-activated cell sorting analysis and used for *in vitro* differentiation in embryoid bodies, which was performed with hanging drop in ESC medium without LIF. After 3 days, embryoid bodies were plated onto gelatine-coated 4-well dishes for another 10 days. The cells were stained with anti-Tuj1 antibody (1:100; Chemicon), anti-α-fetoprotein (AFP) antibody (1:100; R&D Systems), or anti-Flk1 antibody (1:100; R&D Systems).

**Western blot analysis.** Total cell lysates ( $2 \times 10^6$ ) prepared from the ESC and NSC were subjected to western blot analysis for expression of Oct4 (Santa Cruz), Sox2 (Santa Cruz), Klf4 (Abcam) and c-Myc (Abcam). β-actin expression levels in all the samples were used as the loading control (Abcam).

**SSEA-1 and alkaline phosphatase staining.** SSEA-1 and alkaline phosphatase staining was performed with the ES Cell Characterization Kit (Chemicon) according to the manufacturer's protocol.

**Teratoma formation.** iPS cells and NSCs ( $1.5 \times 10^6$  cells per mouse) were injected subcutaneously into the dorsal flank of nude mice. Four weeks after the injection, teratomas that had formed were fixed overnight (approximately 12 h) in 4% PFA and were embedded in paraffin. Sections were stained with haematoxylin and eosin dyes.

**Chimaera formation.** iPS cells were aggregated and cultured with denuded post-compacted 8-cell-stage mouse embryos. In brief, 2-cell-stage embryos were flushed from mice ((C57BL/6 × C3H) F<sub>1</sub> females × CD1 males) at 1.5 d.p.c. and placed in M2 medium and cultured in KSOM medium with 0.1% BSA overnight to the 8-cell stage. Clumps of loosely connected iPS cells (10–20 cells) from short trypsin-treated day-2 cultures were selected and transferred into microdrops of KSOM medium with 10% FCS under mineral oil; each clump was placed in a depression in the microdrop. Meanwhile, batches of 30 to 40 embryos were briefly incubated with acidified Tyrode's solution until the zona pellucida had disintegrated. A single embryo was placed onto the clump. All aggregates were assembled in this manner, and were cultured at 37 °C in an atmosphere of 5% CO<sub>2</sub> in air. After 24 h of culture, most aggregates had formed blastocysts. A total of 64 aggregated blastocysts (2.5 d.p.c.) were transferred into the uterine horns of five pseudopregnant mice (CD-1 background).

25. Szabo, P. E., Hubner, K., Scholer, H. & Mann, J. R. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech. Dev.* 115, 157–160 (2002).
26. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31, e15 (2003).

# The behaviour of *Drosophila* adult hindgut stem cells is controlled by Wnt and Hh signalling

Shigeo Takashima<sup>1</sup>, Marianna Mkrtchyan<sup>1</sup>, Amelia Younossi-Hartenstein<sup>1</sup>, John R. Merriam<sup>1</sup> & Volker Hartenstein<sup>1</sup>

The intestinal tract maintains proper function by replacing aged cells with freshly produced cells that arise from a population of self-renewing intestinal stem cells (ISCs). In the mammalian intestine, ISC self renewal, amplification and differentiation take place along the crypt–villus axis, and are controlled by the Wnt and hedgehog (Hh) signalling pathways<sup>1</sup>. However, little is known about the mechanisms that specify ISCs within the developing intestinal epithelium, or about the signalling centres that help maintain them in their self-renewing stem cell state. Here we show that in adult *Drosophila melanogaster*, ISCs of the posterior intestine (hindgut) are confined to an anterior narrow segment, which we name the hindgut proliferation zone (HPZ). Within the HPZ, self renewal of ISCs, as well as subsequent proliferation and differentiation of ISC descendants, are controlled by locally emanating Wingless (Wg, a *Drosophila* Wnt homologue) and Hh signals. The anteriorly restricted expression of Wg in the HPZ acts as a niche signal that maintains cells in a slow-cycling, self-renewing mode. As cells divide and move posteriorly away from the Wg source, they enter a phase of rapid proliferation. During this phase, Hh signal is required for exiting the cell cycle and the onset of differentiation. The HPZ, with its characteristic proliferation dynamics and signalling properties, is set up during the embryonic phase and becomes active in the larva, where it generates all adult hindgut cells including ISCs. The mechanism and genetic control of cell renewal in the *Drosophila* HPZ exhibits a large degree of similarity with what is seen in the mammalian intestine. Our analysis of the *Drosophila* HPZ provides an insight into the specification and control of stem cells, highlighting the way in which the spatial pattern of signals that promote self renewal, growth and differentiation is set up within a genetically tractable model system.

The hindgut forms the posterior part of the *Drosophila* digestive tract (Fig. 1c). It is composed of large cuboidal epithelial cells (Fig. 1a, e) that secrete cuticle at their apical surface and subserve water absorption. The hindgut epithelium contains a short anterior segment made up of a group of small columnar cells (Fig. 1b, d, f) that differ clearly from posteriorly located hindgut enterocytes (Fig. 1a, e, f, bottom panel). Recent studies showed that a population of small cells within the midgut epithelium acts as self-renewing stem cells for the midgut enterocytes and endocrine cells<sup>2,3</sup>. The presence of small cells in the hindgut led us to examine whether the anterior segment also contains hindgut ISCs. After adult flies were fed 5-bromo-2'-deoxyuridine (BrdU) for 3 to 7 days, BrdU incorporation is found in the anterior segment (Fig. 1g), indicating the presence of proliferative cells in this part of the hindgut, which we will call the hindgut proliferation zone (HPZ). To analyse proliferation in the HPZ further, we induced MARCM clones in adult hindgut<sup>4</sup>. In hindguts fixed 1–2 weeks after clonal induction, clones are small and mostly restricted to the HPZ (Fig. 1h, i). Clones confined to the anterior part of the HPZ range in size from 1 to 6 cells (Fig. 1h). In the

posterior HPZ, clone size was typically larger, between 4 and 12 cells (Fig. 1i). Some clones (4–8 cells in size) were also observed in the differentiated gut (Fig. 1j). Increasing the interval between clonal induction and fixation to 4 weeks resulted in two types of clones that we interpret as stem cell clones and transient clones<sup>3</sup>. The former are large, elongated clones (Fig. 1k) that always include the anterior HPZ (Fig. 1k, right panel) and from there extend posteriorly into the mature hindgut. Transient clones (Fig. 1l) contain 4–12 cells and are most frequently found in the differentiated hindgut, but occasionally also in the posterior part of the HPZ. The clonal data suggest that cells are constantly generated in the HPZ, from where they move posteriorly and become mature hindgut cells.

The HPZ measures approximately 30 cell diameters along the antero-posterior axis. Cells of the first 5–7 rows are elongated transversally; we call this part of the HPZ the spindle cell zone (SCZ). Molecularly, the SCZ is marked by a reporter for Jak/Stat signalling activity, *Stat92E-GFP*<sup>5</sup> (Fig. 1m). The SCZ can be further subdivided into a narrow, anterior zone containing slightly larger *wingless* (*wg*)-positive cells (aSCZ; Fig. 1n), and a wider posterior zone (pSCZ) with *wg*-negative cells. The posterior part of the HPZ consists of 15–20 rows of round cells (round cell zone, RCZ; Fig. 1o), which express *hedgehog* (*hh*).

Whereas proliferating cells are found throughout the HPZ, proliferation rates differ between the SCZ and RCZ. Cells of the SCZ divide less frequently and retain BrdU labelling longer than cells of the RCZ (Fig. 1p–s; see legend for details). The finding suggests a mechanism (Fig. 1t) that decelerates cell-cycle progression within the SCZ, so that cells within this zone behave as slowly cycling ISCs. Progeny of ISCs move posteriorly into the RCZ where they start replicating at a higher rate as they escape the proliferation block that is imposed upon them in the SCZ. Finally, these cells exit the RCZ posteriorly, become post-mitotic and differentiate. The notion of a posterior movement of HPZ cells was supported by a study where two different types of thymidine analogues were applied sequentially (Fig. 1u; see legend for details).

The layout and pattern of proliferation of ISCs within the *Drosophila* hindgut is markedly similar to the way in which ISCs and their progeny are arranged along the crypt–villus axis of the mammalian gut<sup>1</sup>. Little is known about mammalian ISC development, including the time and place at which they and their niche are specified. The *Drosophila* hindgut presents a favourable preparation to address these questions. At an early embryonic stage, the hindgut ectoderm and midgut endoderm emerge as two adjacent lineages<sup>6,7</sup> (Fig. 2a, panel 1). Subsequently, the primordial hindgut cells proliferate and form an elongated epithelial tube that differentiates into the polyploid larval hindgut<sup>7</sup>. At this time a narrow zone of cells that remain diploid<sup>8</sup> and continue to proliferate throughout the larval period can be distinguished at the anterior hindgut boundary (Fig. 2a, panel 2). Called the imaginal ring of the hindgut<sup>9</sup> during the

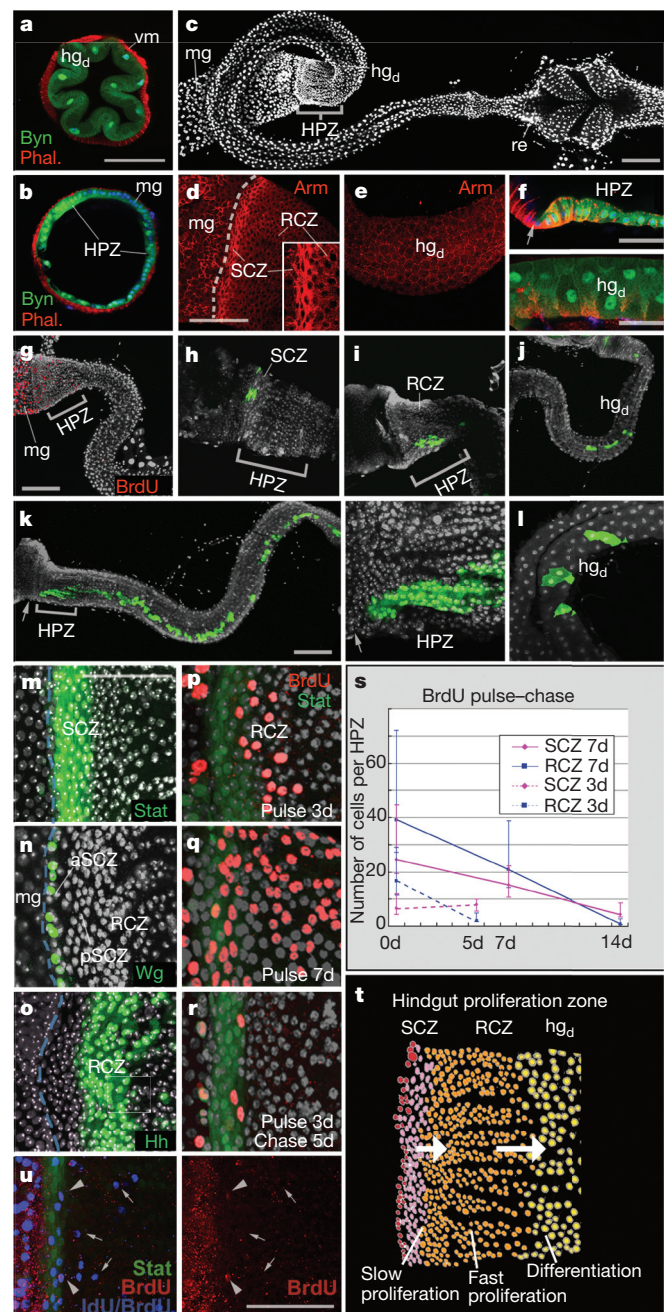
<sup>1</sup>Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, California 90095, USA.



larval stage, these cells eventually develop into the HPZ (Fig. 2a, panels 3–5). The imaginal ring at late embryonic stages (Fig. 2b–d) and in the larva (Fig. 2e–i) is already structured similarly to the adult HPZ, with an aSCZ that expresses both *wg* and *Stat92E-GFP* (Fig. 2b, d, f), a pSCZ that expresses *Stat92E-GFP* (Fig. 2d, f), and an RCZ with high expression levels of *hh* (Fig. 2c, i). Proliferation is found in both the SCZ and RCZ, but is much stronger in the latter (Fig. 2h and Supplementary Fig. 2). Unlike the adult HPZ, cells of the larval imaginal ring/HPZ do not move out posteriorly into the differentiated hindgut. At metamorphosis, cells of the imaginal ring/HPZ expand rapidly over the larval hindgut and differentiate, whereas larval hindgut cells are displaced and undergo cell death (Fig. 2a, panel 4, and Fig. 2j).

In the vertebrate intestine, the canonical Wnt pathway is required for the proliferation and maintenance of ISCs located at the crypt base<sup>10–12</sup>. The Hh signal antagonizes the effect of the Wnt signal, promoting the terminal differentiation of intestinal cells<sup>13,14</sup>, and also controlling the proper spacing of crypts and villi<sup>1,14</sup>. Given that *wg*

and *hh* are expressed in adjacent zones of the *Drosophila* HPZ, the two signals may function in a similar manner in the *Drosophila* hindgut. Armadillo (Arm), the  $\beta$ -catenin homologue in *Drosophila*, is highly accumulated in the cytoplasm and nuclei of the SCZ close to the source of Wg (Fig. 3a, b). This supports the idea that Wg acts as a local signal in the HPZ. To examine the role of the Wg signal further, we overexpressed *wg* and a dominant-negative TCF (*pangolin*), a downstream transcription factor of the Wg signalling pathway, using the Gal4 UAS system<sup>15</sup> combined with *tub-GAL80<sup>ts</sup>* (ref. 16), allowing us temporal control of Gal4 activity. Overexpression of *wg*, starting from early larval stages, results in a marked expansion of the size of the HPZ in late-stage larvae (Fig. 3c, e, g) and in newly eclosed adult flies (Fig. 3d, f, h). A similar effect can be achieved by expressing *wg* exclusively during the adult stage (Fig. 3i, j). The dosage of Wg affects SCZ and RCZ cells differently. At high expression levels (raising larvae at 29 °C, whereby Gal4 UAS is fully active) most cells of the increased HPZ are SCZ cells. They show high levels of cytoplasmic Arm (Supplementary Fig. 2), express *Stat92E-GFP* (Fig. 3g, h) and in many cases have the characteristic shape of spindle cells (Supplementary Fig. 2). On the other hand, when *wg* is weakly expressed (keeping larvae at 25 °C, where Gal4 UAS is less functional), the expansion of *Stat92E-GFP* is less extreme, and one observes a strong increase in the number of RCZ cells (Fig. 3e, f). These data indicate that not only does Wg induce proliferation



**Figure 1 | Proliferating cells of the *Drosophila* hindgut are confined to a morphologically and molecularly defined zone (HPZ).** **a–g**, Structure of the hindgut and its proliferation zone. hg, hindgut; hg<sub>d</sub>, differentiated hindgut cells; HPZ, hindgut proliferation zone; mg, midgut; RCZ, round cell zone; re, rectum; SCZ, spindle cell zone; vm, visceral muscle. **a, b**, Confocal sections of posterior hindgut (**a**) and anterior hindgut (**b**). Visceral muscle layer is labelled with phalloidin (red); nuclei are labelled with TOTO-3 (blue); hindgut epithelial cells express *byn-Gal4>UAS-GFP* (green). **c**, Whole mount of hindgut labelled with nuclear dye Sytox green. **d, e**, Tangential confocal sections of anterior hindgut (**d**, dashed line indicates the midgut–hindgut boundary) and posterior hindgut (**e**) labelled with anti-Arm/ $\beta$ -catenin (red). **f**, Longitudinal confocal sections labelled by the hindgut marker *byn-Gal4>UAS-GFP* (green) and anti-Arm/ $\beta$ -catenin (red). The posterior, differentiated hindgut (**a, e** and bottom panel of **f**) consists of large cuboidal cells surrounded by circular muscle. The HPZ forms a ring at the midgut–hindgut boundary (**c**) and contains small, cylindrical cells (**b**). Membrane-bound Arm outlines the characteristic shape of cells in SCZ, RCZ and differentiated hindgut (**d–f**; inset in **d** shows a high magnification image of SCZ and RCZ). **g**, Proliferative cells that incorporate BrdU (red) are found in the HPZ. **h–l**, MARCM clones (green) induced in adult hindguts and imaged after 2 (**h–j**) or 4 weeks (**k**, stem cell clone; **l**, transient clone). The right panel of **k** is a magnified image of the HPZ. **m–r**, Tangential confocal sections of HPZ. The entire SCZ expresses *Stat92E-GFP* (**m**). Anterior SCZ (aSCZ) expresses Wg (**n**; *wg-Gal4>UAS-mCD8GFP*). The RCZ expresses Hh (**o**; *hh-GAL4>UAS-GFP*). A 3-day pulse of BrdU labels cells predominantly in the RCZ (**p**). A 7-day pulse labels many cells in both SCZ and RCZ (**q**). After a 3-day BrdU pulse followed by a 5-day chase, label is retained in the SCZ, but is all but absent in the RCZ (**r**). **s**, Kinetics of BrdU retention in the HPZ. Dotted lines show 3-day labelling experiment and solid lines show 7-day labelling experiment. A short pulse of BrdU followed by a 5-day chase (dotted lines) does not result in a significant change of the number of BrdU-positive cells in the SCZ (red), but leaves only a few labelled cells (9.5%, compared to 0 days) in the RCZ (blue). If long BrdU pulses are followed by chase periods, the number of labelled cells declines steadily for both SCZ and RCZ, but the decline is much steeper for the RCZ than the SCZ (compare solid blue and red lines). Error bars indicate s.d. **t**, Schematic illustration of the HPZ. **u**, Double labelling of proliferating cells of the HPZ with a 5-day IdU pulse followed by a 3-day BrdU pulse, showing posterior movement of cells. A BrdU-specific antibody (red) and an antibody that recognizes both IdU and BrdU (blue) were used. Arrows indicate several IdU-positive cells which have moved posteriorly; arrowheads indicate BrdU-positive cells still confined to the anterior HPZ. Anterior is to left in all pictures and in all figures except Fig. 1a, b (cross sections). Scale bars: **a**, 100  $\mu$ m (also for **b**); **c**, 100  $\mu$ m; **d**, 100  $\mu$ m (also for **e, h, i**, right panel of **k**, and **l**); **f**, 25  $\mu$ m; **g**, 100  $\mu$ m (also for **j** and left panel of **k**); **m**, 100  $\mu$ m (also for **n–r**); **u**, 100  $\mu$ m.

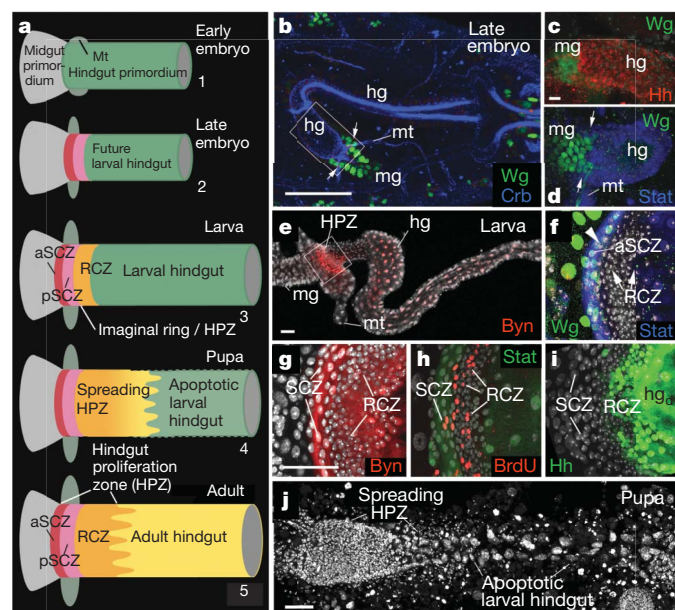


within the HPZ, but at high levels of expression it causes HPZ cells to remain in the SCZ cell state as self-renewing ISCs. Such a function of *wg* would explain how the size of the SCZ remains stable after many rounds of cell division that occur in the HPZ. Cells in and adjacent to the Wg-expressing aSCZ receive high concentrations of the Wg signal and express the ISC fate; as the ISC population gradually grows, cells farthest away from the Wg source escape from the effective range of Wg and switch their fate to that of fast cycling hindgut progenitors that form the RCZ. This model implies that the aSCZ forms part of the niche that maintains the self-renewing properties of *Drosophila* hindgut ISCs. Notably, the Wg-positive cells of the aSCZ/niche clearly form part of the stem cell pool that incorporates BrdU (Fig. 3k).

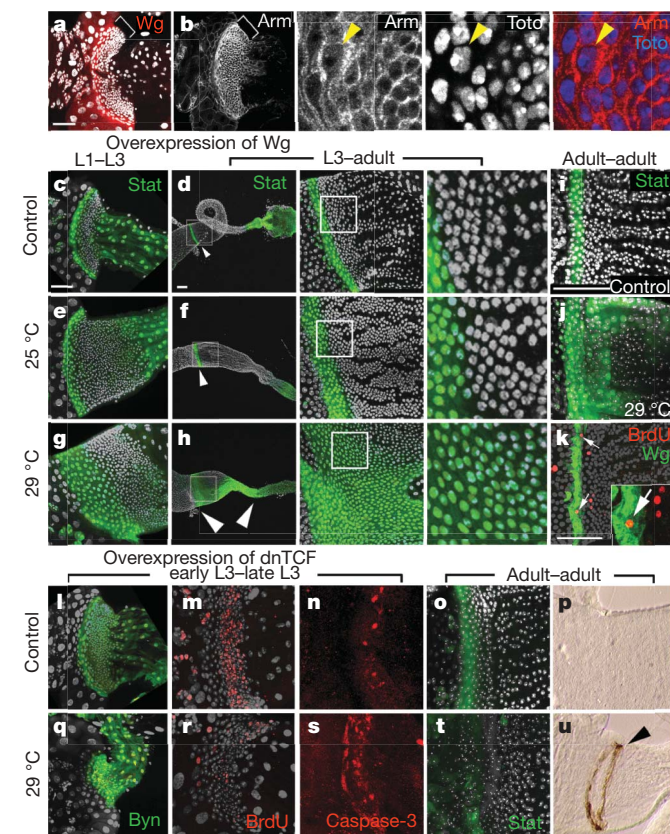
Overexpression of dominant-negative *TCF* (*dnTCF*)<sup>17</sup> results in a strong reduction of the HPZ in late third instar larvae (Fig. 3l, q). If these animals undergo metamorphosis and eclose, the hindgut is missing entirely (data not shown). Applying the BrdU incorporation assay in *dnTCF*-overexpressing larvae revealed a decrease in

proliferation in the HPZ (Fig. 3m, r). At the same time, strongly increased numbers of Caspase-3-positive cells in the HPZ revealed an increase in apoptosis (Fig. 3n, s). If *dnTCF* overexpression is restricted to the adult phase, we observe a similar phenotype (Fig. 3o, p, t, u). These data further support the notion that *wg* is required for proliferation and maintenance of the HPZ in larval and adult life periods.

To address the role of the Hh signalling pathway in the *Drosophila* hindgut we overexpressed a repressor isoform of *cubitus interruptus*



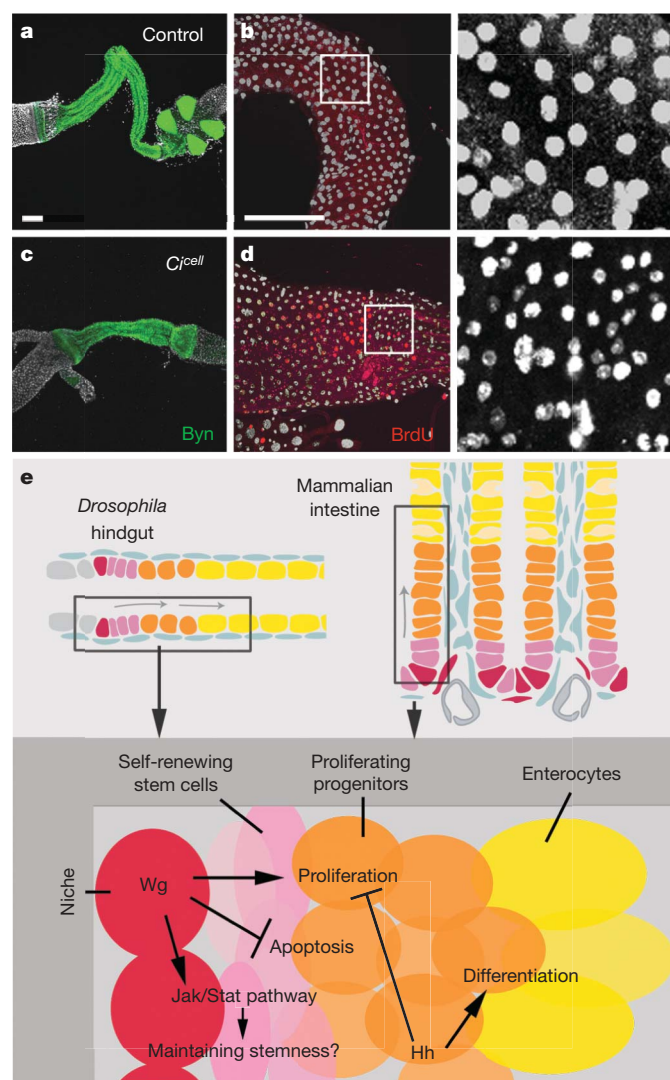
**Figure 2 | The HPZ is formed in the embryo and produces the adult hindgut during the larval and pupal phases.** **a**, Schematic illustration of hindgut development from embryo to adult. In the early embryo, the primordium of the hindgut forms a proliferating epithelial tube (1). Towards the end of embryogenesis (2), most cells differentiate into the larval hindgut; cells at the anterior hindgut boundary give rise to the HPZ, which proliferates in the larva (3). During this stage, a clear zonation into aSCZ, pSCZ and RCZ becomes apparent. During metamorphosis (4) the larval hindgut undergoes apoptosis and is replaced by the spreading HPZ. Only the anterior cells retain their proliferative activity and become the HPZ of the adult (5). **b–d**, Origin of the HPZ in the late embryo. hg, hindgut; mg, midgut; mt, Malpighian tubules. A membrane protein Crumbs (Crb, blue in **b**) demarcates hindgut. A narrow zone of Wg expression (green) straddles the boundary between midgut (Crb-negative) and hindgut (**b**). The boxed area indicates a segment of hindgut shown in **c**, **d**. Hh is expressed adjacent and posterior to Wg (**c**). *Stat92E-GFP* expression (blue in **d**) includes the Hh and posterior part of the Wg-positive domain. The midgut–hindgut boundary is demarcated by arrows. **e–i**, The larval HPZ. Small, proliferating cells form the HPZ at the anterior boundary of the hindgut (**e**). The boxed area indicates a segment of hindgut shown in **f–i**. Structurally, the larval HPZ is divided into an SCZ and an RCZ (**g**). The rate of proliferation is higher in the RCZ than in the SCZ (**h**). Thus, a 4-h pulse of BrdU labelled 3% of nuclei in the SCZ (marked by *Stat92E-GFP*) compared to 33% in the RCZ (see also Supplementary Fig. 1). The anterior SCZ (aSCZ) expresses Wg and the entire SCZ expresses *Stat92E-GFP* (**f**); the RCZ expresses Hh (**i**). **j**, Hindgut at 24 h after puparium formation (metamorphosis). All nuclei are labelled by Sytox green. The HPZ spreads posteriorly to replace apoptotic larval hindgut cells. Scale bars: **b**, 50  $\mu$ m; **c**, 10  $\mu$ m (also for **d**); **e**, 50  $\mu$ m; **g**, 50  $\mu$ m (also for **f**, **h**, **i**); **j**, 50  $\mu$ m.



**Figure 3 | Wg signalling promotes proliferation and maintenance of the hindgut stem cell population.** **a**, **b**, Arm/ $\beta$ -catenin is accumulated at high levels in the SCZ. The entire HPZ in the larval hindgut is indicated by a bracket. The localization of Arm protein becomes cytoplasmic/nuclear (**b**); the second to fourth panels are a magnification ( $\times 8.0$ ) of the first panel in the cells close to the Wg-expressing cells (**a**). Arrowheads in **b** indicate an SCZ cell. **c–j**, Overexpression of *wg* causes an expansion of the SCZ, assayed by *Stat92E-GFP* (green). Labels along the top indicate the developmental time interval at which *byn-Gal4*-driven *wg* was active (L1, first instar; L3, third instar larva). Middle panels of **d**, **f**, **h** show magnified views ( $\times 5.0$ ) of the HPZ (boxed area in left panels); right panels show further magnifications ( $\times 3.5$ ) of the boxed area in the middle panels. Moderate (raising flies at 25  $^{\circ}$ C; **e**, **f**) and strong (29  $^{\circ}$ C; **g**, **h**) overexpression of *wg* is shown. Arrowheads in **d**, **f** and **h** indicate the expression domain of *Stat92E-GFP*. Overexpression of *wg* after eclosion also causes widening of the *Stat92E-GFP*-positive SCZ (**j**). **k**, *wg*-expressing aSCZ cell in wild-type hindgut (green) incorporates BrdU (red) after a 3-day pulse. **l–u**, *byn-Gal4*-driven expression of *dnTCF* as an inhibitor of Wg signalling. Labels along the top indicate the time period of temperature shifts that caused overexpression of *dnTCF*. **l–p**, Wild-type controls. In the larval period, *dnTCF* leads to a decrease in size of the HPZ (**q**), caused by reduction of proliferation (**r**), as well as increased cell death (**s**). When overexpressed in the adult, *dnTCF* causes reduction of the SCZ as monitored by expression of *Stat92E-GFP* (**t**), and abnormal sclerotization of the HPZ resulting from increased cell death (**u**, arrowhead; panels **p** and **u** show the anterior hindgut of wild-type and *dnTCF* flies, respectively, in interference contrast). Scale bars: **a**, 50  $\mu$ m (also for first panel of **b**); **c**, 50  $\mu$ m (also for **d**, **f** and **h** (middle panels), **e**, **g**, **i**, **p**, **q**, **u**); **d**, 100  $\mu$ m (also for **f**, **h** (left panels)); **i**, 50  $\mu$ m (also for **j**); **k**, 50  $\mu$ m (also for **m–o**, **r–t**).

( $Ci^{cell}$ )<sup>18</sup>, a transcription factor acting downstream of the Hh signalling pathway. Continuous expression of  $Ci^{cell}$  from larval stages into the adult results in a markedly shortened hindgut (Fig. 4a, c) that consisted almost entirely of small, undifferentiated cells normally found only in the HPZ (Fig. 4b, d, right panels).  $Ci^{cell}$ -expressing flies actively incorporate BrdU all over their rudimentary hindgut (Fig. 4b, d), indicating that by inhibiting the Hh pathway, hindgut cells are prevented from leaving the cycle of proliferation and from differentiating.

Our data show that *Drosophila* hindgut stem cells are controlled by the Wg and Hh signalling pathways in a way that could be fundamentally similar to what is seen in the mammalian intestine (Fig. 4e). In both systems, the Wnt/Wg signal acts in/through the niche, where it promotes slow proliferation and maintains a population of ISCs. Meanwhile, the Hh pathway promotes differentiation and sets a limit on the size of the proliferating cell pool. These



**Figure 4 | The Hh signal promotes differentiation of the *Drosophila* hindgut epithelium.** **a–d**, Following an overexpression of  $Ci^{cell}$ , the hindgut (labelled by  $byn-Gal4>UAS-GFP$ ) becomes short (**c**) compared to wild-type control (**a**). Posterior hindgut cells are non-proliferating (**b**) and large (right panel, which is a magnification ( $\times 4.0$ ) of boxed area in left panel) in the control, whereas they incorporate BrdU (**d**) and are small (right panel, which is a magnification ( $\times 4.0$ ) of boxed area in left panel) in  $Ci^{cell}$ -overexpressing hindgut. **e**, Schematic diagram summarizing the functions of Wg and Hh pathways in the HPZ. Top panel: a comparison between the *Drosophila* hindgut and individual crypts/villi of the mammalian intestine. Red, cells transmitting niche signals; magenta and orange, stem cells and proliferative progenitor cells, respectively; yellow, differentiated enterocytes. Scale bars: **a**, 100  $\mu$ m (also for **c**); **b**, 100  $\mu$ m (also for **d**).

remarkable similarities suggest a true homology between the vertebrate and invertebrate ISC system, supporting the rationale of using *Drosophila* to explore further the genes involved in the development of ISCs. Aside from Wnt and Hh signals, several other pathways, including the BMP and Notch pathways, have a role in intestinal development and maintenance. Notch seems to act downstream of Wnt, and controls a switch between the different cell fates (secretory versus absorptive cells), but it also acts in conjunction with Wnt to control the stem cell compartment<sup>1</sup>. Notch has a similar role in the fly midgut<sup>2,3,19</sup>. Notch activity is observed in the adult *Drosophila* hindgut (S.T. and V.H., unpublished), and future studies will clarify how this pathway interacts with Hh and Wnt signalling. The *Drosophila* system offers the advantage of considerably less complexity than its mammalian counterpart, in terms of cell number and different cell types. Thus, in the mammalian intestine, both ISCs and proliferating ISC descendants are found in the crypt epithelium, which also expresses both ligands and receptors of the Wnt and Hh pathways<sup>13,14,20</sup>. Furthermore, the mesenchyme surrounding the crypt forms an integral part of the signalling mechanism<sup>1</sup>. It is not yet clear at single-cell resolution how emission and reception of the different signals acting together in the crypt relate to each other. By contrast, the single-layered hindgut architecture, as well as the linear arrangement of both cell types and signals in the HPZ, may make it easier to conceptualize possible roles of signalling pathways, or to interpret the outcome of experimental manipulations.

## METHODS SUMMARY

**Fly strains.** Flies were reared with normal fly food at room temperature or in incubators at 18 °C, 25 °C, or 29 °C. The strains used in this study are listed in the Methods.

**Temporal overexpression of genes and MARCM clone assay.** The genotypes of flies used for temporal overexpression of genes were as follows:  $UAS-wg/+$ ;  $tub-GAL80^{ts}/byn-Gal4$   $UAS-GFP$ ,  $tub-GAL80^{ts}/+$ ;  $UAS-dnTCF/byn-Gal4$   $UAS-GFP$ ,  $tub-GAL80^{ts}/+$ ;  $UAS-Ci^{cell}/byn-Gal4$   $UAS-GFP$ ,  $tub-GAL80^{ts}/+$ ;  $Stat92E-GFP/UAS-wg$ ;  $byn-Gal4$   $UAS-lacZ/+$ ,  $tub-GAL80^{ts}/+$ ;  $Stat92E-GFP/+$ ;  $UAS-dnTCF/byn-Gal4$   $UAS-lacZ$ . The genotype for MARCM clone induction was:  $hs\ flp$  FRT19A  $tub-GAL80/FRT19A$ ;  $byn-Gal4$   $UAS-GFP/+$ . Heat-shock was given for 1 h at 37 °C.

**Antibodies.** Antibodies used in this study were: mouse anti-Arm (1:10, DSHB, University of Iowa), mouse anti- $\beta$ -gal (1:100, Promega), rabbit anti- $\beta$ -gal (1:1,000, Cappel), mouse anti-BrdU/IdU (clone B44, 1:100, BD Biosciences), rat anti-BrdU (clone BU1/75 (ICR1), abcam), rabbit anti-Caspase-3 (1:100, Cell Signaling Technology), mouse anti-Crumbs (1:40, DSHB), mouse anti-GFP (1:100, Sigma), and mouse anti-Wg (1:50, DSHB).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 20 February; accepted 9 June 2008.

Published online 16 July 2008.

- Crosnier, C., Stamatakis, D. & Lewis, J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nature Rev. Genet.* **7**, 349–359 (2006).
- Micchelli, C. A. & Perrimon, N. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* **439**, 475–479 (2006).
- Ohlstein, B. & Spradling, A. The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature* **439**, 470–474 (2006).
- Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development. *Trends Neurosci.* **24**, 251–254 (2001).
- Bach, E. A. *et al.* GFP reporters detect the activation of the *Drosophila* JAK/STAT pathway in vivo. *Gene Expr. Patterns* **7**, 323–331 (2007).
- Murakami, R., Takashima, S. & Hamaguchi, T. Developmental genetics of the *Drosophila* gut: specification of primordia, subdivision and overt-differentiation. *Cell. Mol. Biol.* **45**, 661–676 (1999).
- Lengyel, J. A. & Iwaki, D. D. It takes guts: the *Drosophila* hindgut as a model system for organogenesis. *Dev. Biol.* **243**, 1–19 (2002).
- Smith, A. V. & Orr-Weaver, T. L. The regulation of the cell cycle during *Drosophila* embryogenesis: the transition to polyteny. *Development* **112**, 997–1008 (1991).
- Robertson, C. W. The metamorphosis of *Drosophila melanogaster*, including an accurately timed account of the principal morphological changes. *J. Morphol.* **59**, 351–399 (1936).
- Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19**, 379–383 (1998).



11. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* **17**, 1709–1713 (2003).
12. Kuhnert, F. *et al.* Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc. Natl Acad. Sci. USA* **101**, 266–271 (2004).
13. van den Brink, G. R. *et al.* Indian Hedgehog is an antagonist of Wnt signaling in colonic epithelial cell differentiation. *Nature Genet.* **36**, 277–282 (2004).
14. Madison, B. B. *et al.* Epithelial hedgehog signals pattern the intestinal crypt-villus axis. *Development* **132**, 279–289 (2005).
15. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
16. McGuire, S. E., Le, P. T., Osborn, A. J., Matsumoto, K. & Davis, R. L. Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302**, 1765–1768 (2003).
17. van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene dTCF. *Cell* **88**, 789–799 (1997).
18. Methot, N. & Basler, K. Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. *Cell* **96**, 819–831 (1999).
19. Ohlstein, B. & Spradling, A. Multipotent *Drosophila* intestinal stem cells specify daughter cell fates by differential notch signaling. *Science* **315**, 988–992 (2007).
20. Gregorieff, A. *et al.* Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* **129**, 626–638 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank DSHB, NIG-Fly and Bloomington stock center for flies and antibodies. We also thank all members of the Hartenstein laboratory and Merriam laboratory for discussions. This work was supported by an NIH grant to J.R.M. and V.H.

**Author Contributions** S.T., J.R.M. and V.H. designed the project. The experiments were carried out by S.T., M.M. and A.Y.-H.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to V.H. ([volkerh@mcdb.ucla.edu](mailto:volkerh@mcdb.ucla.edu)).



## METHODS

**Fly strains.** Flies used in this study were (donors in parentheses): *byn<sup>apro</sup>/TM3* (R. Murakami), *byn-Gal4* (J. Lengyel), *hh-Gal4* (K. Basler), *wg-Gal4* (H. Krause), *UAS-C<sup>cell</sup>* (K. Basler), and the following stocks were obtained from Bloomington stock centre: *OregonR<sup>w1118</sup>*, *hh-lacZ<sup>p30</sup>*, *UAS-arm<sup>S10</sup>*, *UAS-EGFP*, *UAS-mCD8GFP*, *UAS-dnTCF/pangolin*, *UAS-wg*, *tub-GAL80<sup>LL1</sup>*, *tub-GAL80<sup>ts</sup> 7*, *tub-GAL80<sup>ts</sup> 20*, or from the National Institute of Genetics, Japan: *wg-lacZ/CyO*, *UAS-lacZ*. All flies were reared with normal fly food at room temperature or in incubators at 18 °C, 25 °C, or 29 °C.

**BrdU incorporation assay.** Labelling proliferative cells with BrdU was performed by feeding adult flies or larvae with BrdU. Adult flies of 1 week after eclosion or older were reared with normal fly food containing BrdU (1 mg ml<sup>-1</sup>, Sigma) for 3 or 7 days consecutively. Third instar larvae were fed with the same food containing BrdU for 2–4 h. Hindguts of adults or larvae were dissected and fixed immediately after BrdU labelling or subsequently reared with normal fly food without BrdU for another 7–14 days before dissection, then treated with 2 N HCl for 30 min on ice, before being processed for antibody staining. IdU (Chem-Impex International) incorporation was performed as above.

**Temporal overexpression of genes and MARCM clone assay.** *tubulin (tub)-GAL80<sup>ts</sup>* was combined with the conventional Gal4 UAS system. Larvae and adult flies, which were raised at 18 °C, were shifted to 29 °C (or 25 °C) to inactivate *GAL80<sup>ts</sup>*, which inhibits the Gal4-driven expression of genes following the UAS. The genotypes of flies used in this experiment were as follows: *UAS-wg/+*; *tub-GAL80<sup>ts</sup>/byn-Gal4 UAS-GFP*, *tub-GAL80<sup>ts</sup>/+*; *UAS-dnTCF/byn-Gal4 UAS-GFP*, *tub-GAL80<sup>ts</sup>/+*; *UAS-C<sup>cell</sup>/byn-Gal4 UAS-GFP*, *tub-GAL80<sup>ts</sup>/+*; *Stat92E-GFP/UAS-wg*; *byn-Gal4 UAS-lacZ/+*, *tub-GAL80<sup>ts</sup>/+*; *Stat92E-GFP/+*; *UAS-dnTCF/byn-Gal4 UAS-lacZ*. For MARCM clone induction, female flies carrying *hs flp FRT19A tub-GAL80/FRT19A*; *byn-Gal4 UAS-GFP/+* were heat-shocked for 1 h at 37 °C. *byn-Gal4* is expressed exclusively in hindgut throughout embryonic and post-embryonic development and therefore directed GFP expression specifically to hindgut clones.

**Antibody staining.** Hindguts were fixed with 4% formaldehyde/PBT for 45 min at room temperature then treated with primary antibody overnight at 4 °C. After treatment with secondary antibody conjugated with fluorescent dye, hindguts were mounted with Vectashield (Sigma). A nuclear indicator (TOTO-3) was added to the mounting medium if necessary (1:2,000 dilutions). Antibodies used in this study were: mouse anti-Arm (1:10, DSHB, University of Iowa), mouse anti-β-gal (1:100, Promega), rabbit anti-β-gal (1:1,000, Cappel), mouse anti-BrdU/IdU (clone B44, 1:100, BD Biosciences), rat anti-BrdU (clone BU1/75 (ICR1), abcam), rabbit anti-Caspase-3 (1:100, Cell Signaling Technology), mouse anti-Crumbs (1:40, DSHB), mouse anti-GFP (1:100, Sigma), and mouse anti-Wg (1:50, DSHB).

## LETTERS

# Blocking VEGFR-3 suppresses angiogenic sprouting and vascular network formation

Tuomas Tammela<sup>1</sup>, Georgia Zarkada<sup>1</sup>, Elisabet Wallgard<sup>2\*</sup>, Aino Murto<sup>1\*</sup>, Steven Suchting<sup>3</sup>, Maria Wirzenius<sup>1</sup>, Marika Waltari<sup>1</sup>, Mats Hellström<sup>2</sup>, Tibor Schomber<sup>4</sup>, Reetta Peltonen<sup>5</sup>, Catarina Freitas<sup>3</sup>, Antonio Duarte<sup>6</sup>, Helena Isoniemi<sup>5</sup>, Pirjo Laakkonen<sup>1</sup>, Gerhard Christofori<sup>4</sup>, Seppo Ylä-Herttuala<sup>7</sup>, Masabumi Shibuya<sup>8</sup>, Bronislaw Pytowski<sup>9</sup>, Anne Eichmann<sup>3</sup>, Christer Betsholtz<sup>2</sup> & Kari Alitalo<sup>1</sup>

Angiogenesis, the growth of new blood vessels from pre-existing vasculature, is a key process in several pathological conditions, including tumour growth and age-related macular degeneration<sup>1</sup>. Vascular endothelial growth factors (VEGFs) stimulate angiogenesis and lymphangiogenesis by activating VEGF receptor (VEGFR) tyrosine kinases in endothelial cells<sup>2</sup>. VEGFR-3 (also known as FLT-4) is present in all endothelia during development, and in the adult it becomes restricted to the lymphatic endothelium<sup>3</sup>. However, VEGFR-3 is upregulated in the microvasculature of tumours and wounds<sup>4,5</sup>. Here we demonstrate that VEGFR-3 is highly expressed in angiogenic sprouts, and genetic targeting of VEGFR-3 or blocking of VEGFR-3 signalling with monoclonal antibodies results in decreased sprouting, vascular density, vessel branching and endothelial cell proliferation in mouse angiogenesis models. Stimulation of VEGFR-3 augmented VEGF-induced angiogenesis and sustained angiogenesis even in the presence of VEGFR-2 (also known as KDR or FLK-1) inhibitors, whereas antibodies against VEGFR-3 and VEGFR-2 in combination resulted in additive inhibition of angiogenesis and tumour growth. Furthermore, genetic or pharmacological disruption of the Notch signalling pathway led to widespread endothelial VEGFR-3 expression and excessive sprouting, which was inhibited by blocking VEGFR-3 signals. Our results implicate VEGFR-3 as a regulator of vascular network formation. Targeting VEGFR-3 may provide additional efficacy for anti-angiogenic therapies, especially towards vessels that are resistant to VEGF or VEGFR-2 inhibitors.

During late embryogenesis and in the adult, blood vessels form primarily by angiogenesis—that is, sprouting from pre-existing vessels. VEGF potently promotes angiogenesis and is indispensable for vascular development<sup>6,7</sup>, whereas VEGFR-2 tyrosine kinase is the primary receptor transmitting VEGF signals in endothelial cells<sup>8,9</sup>. Angiogenic sprouting involves specification of subpopulations of endothelial cells into tip cells that respond to VEGF guidance cues and into stalk cells that follow the tip cells and proliferate to form the vascular network<sup>10</sup>. Recent evidence indicates that VEGF induces the Notch ligand Delta-like 4 (DLL4) in the tip cells, which leads to suppression of excess sprouts in adjacent endothelial cells<sup>11–17</sup>.

VEGFR-3 is activated by VEGF-C and VEGF-D (also known as FIGF), which can also stimulate VEGFR-2 after proteolytic processing (reviewed in ref. 18). Furthermore, VEGF-C has been shown to induce

the formation and activation of VEGFR-2–VEGFR-3 heterodimers<sup>19</sup>. *Vegfr3*-gene-targeted mice exhibit marked defects in arterial–venous remodelling of the primary vascular plexus, resulting in lethality by embryonic day (E)10.5 (ref. 20). Correspondingly, knockdown of the *Vegfr3* homologue in zebrafish results in defective segmental artery morphogenesis<sup>21</sup>. We recently observed that blocking antibodies against VEGFR-3 inhibit tumour growth by inhibiting angiogenesis<sup>22</sup>. These results suggested that VEGFR-3 signalling is required for both developmental and tumour angiogenesis, raising the question of how VEGFR-3 regulates this process.

We validated vascular VEGFR-3 expression in tumours by growing syngeneic tumours in *Vegfr3*<sup>+/LacZ</sup> mice, which allows visualization of *Vegfr3*-expressing cells using the  $\beta$ -galactosidase reporter driven by the *Vegfr3* promoter. Blood vessels within the tumour tissue showed *Vegfr3* promoter activity in grafted B16 melanomas (Fig. 1a, arrows) and Lewis lung carcinomas (data not shown), but not in the adjacent peritoneum. We also detected blood vascular VEGFR-3 expression in human colon carcinoma metastases of the liver, in all three human tumour xenografts analysed, and in transgenic Rip1Tag2 mouse insulinomas (Supplementary Fig. 1). Notably, VEGFR-3 immunostaining in the tumours was prominent in endothelial sprouts (Supplementary Fig. 1c–e). Angiogenic vessels positive for VEGFR-3 but negative for the lymphatic vessel marker LYVE-1 were also found in developing ovarian follicles of superovulated mice (Supplementary Fig. 2). These data indicated that VEGFR-3 is expressed specifically in blood vessels undergoing angiogenesis.

Because blood vessels are disorganized in tumour tissues, we studied VEGFR-3 expression in the mouse retina, where new vessels form in an organized and directional manner during the postnatal period by sprouting angiogenesis that proceeds centrifugally from the optic disc<sup>10</sup>. VEGFR-3 was strongly expressed in the angiogenic vessel front during the early postnatal period (days 1–7; Fig. 1b and Supplementary Fig. 3), but not in mature vessels at postnatal day (P)28 (Supplementary Fig. 3b). VEGFR-3 was also weakly expressed in the veins and venules (Fig. 1b, c and Supplementary Fig. 3), reflecting the pattern observed during early vascular development<sup>20</sup>. High-resolution analysis revealed that the  $\beta$ -galactosidase reporter driven by the VEGFR-3 promoter localized to the leading tip cells of endothelial sprouts (Fig. 1c and Supplementary Fig. 3c), whereas immunofluorescence analysis by confocal microscopy demonstrated

<sup>1</sup>Molecular/Cancer Biology Laboratory and Ludwig Institute for Cancer Research, Biomedicum Helsinki and the Haartman Institute University of Helsinki, PO Box 63 (Haartmaninkatu 8), 00014 Helsinki, Finland. <sup>2</sup>Division of Matrix Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden. <sup>3</sup>Institut National de la Santé et de la Recherche Médicale U833, Collège de France, 11 Place Marcelin Berthelot, 75005 Paris, France. <sup>4</sup>Center of Biomedicine, Department of Clinical-Biological Sciences, University of Basel, Mattenstrasse 28, CH-4058 Basel, Switzerland. <sup>5</sup>Department of Transplantation and Hepatic Surgery, Helsinki University Central Hospital, PO Box 263, 00029 Helsinki, Finland. <sup>6</sup>The Interdisciplinary Centre of Research in Animal Health (CIISA), Faculty of Veterinary Medicine, Technical University of Lisbon, 1300-474 Lisbon, Portugal. <sup>7</sup>A. I. Virtanen Institute and Department of Medicine, University of Kuopio, PO Box 1627, 70211 Kuopio, Finland. <sup>8</sup>Department of Molecular Oncology, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo 113-8519, Japan. <sup>9</sup>ImClone Systems, 180 Varick Street, New York 10014, USA.

\*These authors contributed equally to this work.

that VEGFR-3 is present in the filopodial extensions of the tip cells (Fig. 1d), where previous studies have reported prominent VEGFR-2 expression<sup>10,16</sup>. We detected VEGF-C protein and promoter activity in leukocytes that were positioned in the region of vascular network formation (Supplementary Fig. 4a–e). VEGF-C bound to endothelial cells was detected by immunostaining (Supplementary Fig. 4a, b).

In developing embryos, the most intense VEGFR-3 staining localized to areas of active angiogenesis at E8.5–E9.5 (Fig. 1e, arrowheads). High-resolution analysis by immunofluorescence indicated intense expression in actively sprouting intersomitic vessels, whereas few sprouts and low levels of VEGFR-3 were detected in the adjacent dorsal aorta at E9.25 (Fig. 1f). VEGF-C expression was detected in the endothelium of actively growing intersomitic vessels at E9.0 and E9.5

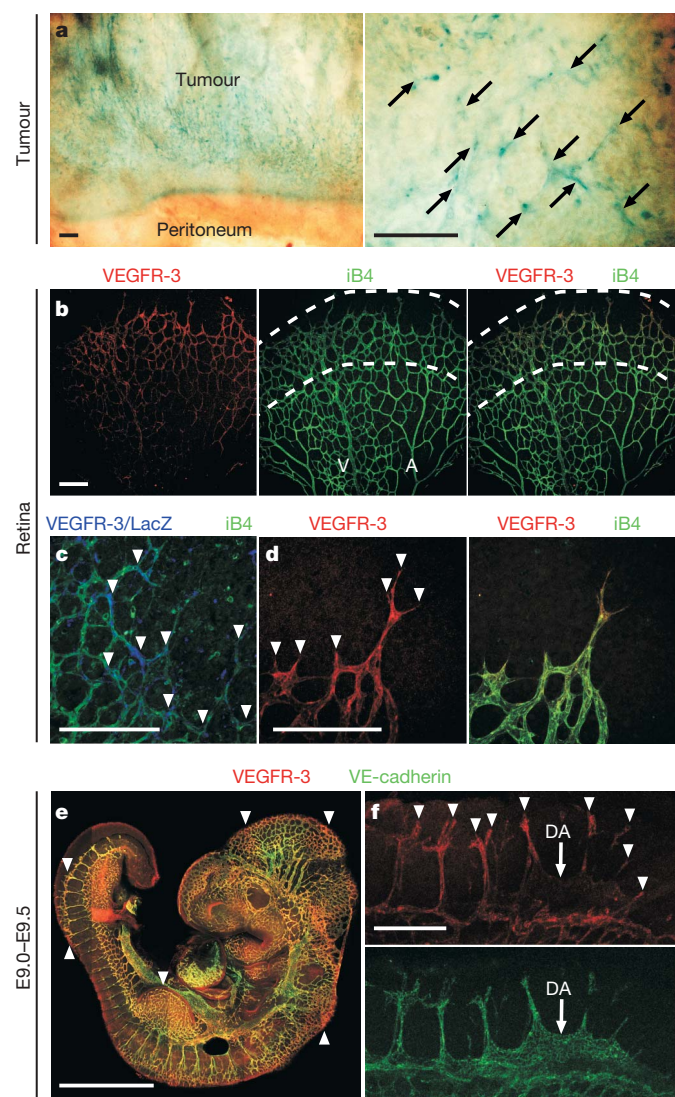
in *Vegfr3*<sup>+LacZ</sup> embryos (Supplementary Fig. 4f–i and data not shown). Interestingly, the vasculature of the head appeared less complex at E8.5 (Supplementary Fig. 5), and intersomitic vessels were smaller and had fewer sprouts at E9.0 in *Vegfr3*-gene-targeted embryos compared to wild-type littermates (Supplementary Fig. 6). VEGFR-2 expression was not affected by loss of *Vegfr3* (Supplementary Fig. 7), indicating that VEGFR-3 signalling is directly required for vascular development.

To inhibit VEGFR-3 signalling in the postnatal retina, newborn pups were given monoclonal antibodies that specifically block ligand binding to VEGFR-3. Antibodies that block VEGFR-2 were used as a positive control for angiogenesis inhibition, and non-specific rat immunoglobulin G (IgG) as a negative control. The vascular network appeared less dense in retinas of pups that were given anti-VEGFR-3 antibodies, whereas only rudimentary vessels were found in pups injected with anti-VEGFR-2 antibodies (Fig. 2a). Interestingly, when both VEGFR-2 and VEGFR-3 antibodies were administered in combination, a more marked reduction of vascular surface area and endothelial sprouts was obtained (Fig. 2b–e), suggesting that VEGFR-3 signalling in part compensates for lack of VEGFR-2 signalling. Notably, blocking VEGFR-3 appeared to most markedly affect the number of sprouts and branching points (Fig. 2c, d). However, we also detected a decrease in the number of proliferating endothelial cells in mice that received VEGFR-3-blocking antibodies (Fig. 2e), which is in line with findings indicating that VEGFR-3 signals promote the proliferation and survival of endothelial cells in culture<sup>23,24</sup>.

VEGFR-3 antibodies or an adenovirally delivered soluble VEGFR-3–Ig fusion protein<sup>25</sup> also markedly reduced the number of VEGFR-3-positive endothelial sprouts in lung carcinoma xenografts (Fig. 2f, g and data not shown), suggesting that VEGFR-3 inhibitors inhibit angiogenesis by suppressing endothelial sprouting. The combination of VEGFR-3 and VEGFR-2 antibodies for 6 days led to a significant decrease in blood vessel density in LNM35 and B16 tumours when compared to administration of either antibody alone (Fig. 2h and not shown). Importantly, the combination of VEGFR-3 and VEGFR-2 antibodies reduced vascular surface area 45%, viable tumour area 32% and tumour growth 41% more effectively than blockade of VEGFR-2 alone (Fig. 2j–l).

To elucidate how VEGFR-3 expression is induced during angiogenesis, we used a model with less confounding factors than in tumours. Expression of human VEGF or VEGF-E—a VEGFR-2-specific ligand—in the skin under the control of the keratin 14 promoter (*K14*-VEGF or *K14*-VEGF-E, respectively, induced marked blood vessel hyperplasia and upregulation of VEGFR-3 in cutaneous blood vessels, whereas the vessels were VEGFR-3-negative in wild-type littermates (Fig. 3a and Supplementary Fig. 8a). Surprisingly, both vascular phenotypes were greatly attenuated in *K14*-VEGF;*K14*-VEGFR-3-Ig and *K14*-VEGF-E;*K14*-VEGFR-3-Ig compound transgenic mice (Fig. 3a, b and Supplementary Fig. 8b, c), indicating that VEGFR-3 signals contributed to the excess angiogenesis. In contrast, the blood vessels of *K14*-VEGFR-3-Ig single-transgenic mice appeared normal, as previously published (Fig. 3a, b)<sup>25</sup>. These results suggest that the angiogenic VEGFR-3 signal is active predominantly during settings of angiogenic invasion of tissues, such as during early embryonic development, in the postnatal retina, and in tumours.

To dissect the VEGFR-2 and VEGFR-3 pathways *in vivo*, we used adenoviral gene transfer vectors in the mouse ear skin for overexpression of VEGFs. When VEGF or VEGF-E was expressed together with mouse VEGF-D—a VEGFR-3-specific ligand<sup>26</sup>—we observed a significant increase in angiogenesis when compared to expression of these factors alone (Fig. 3c, Supplementary Fig. 9). However, mouse VEGF-D was not able to induce angiogenesis on its own, suggesting that blood vessels need to be primed by VEGFR-2 signals to become responsive to VEGFR-3 ligands (Fig. 3c and Supplementary Fig. 9). In a reversed experiment, we overexpressed both VEGF and VEGF-C, which are both commonly expressed in tumours, in the ear skin.

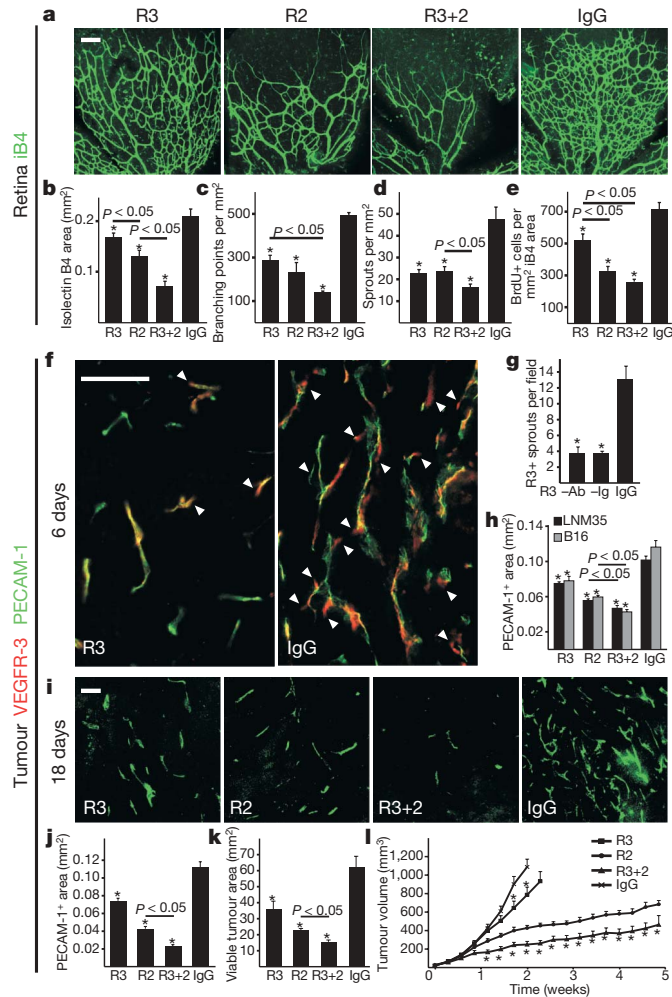


**Figure 1 | VEGFR-3 is expressed in the tumour vasculature and localizes to endothelial tip cells.** **a**, VEGFR-3 expression (arrows, blue) detected by X-gal staining in a syngeneic B16 melanoma implanted subcutaneously into a *Vegfr3*<sup>+LacZ</sup> heterozygous mouse. The peritoneal membrane, which contains normal blood vessels, has no staining. **b**, VEGFR-3 staining (red) of a mouse retina at P5. Blood vessels are visualized with isolectin B4 staining (iB4, green). A, artery; V, vein. **c**, A single confocal section showing VEGFR-3 (blue) after X-gal staining, and isolectin B4 (green) in the angiogenic front at P5. **d**, VEGFR-3 (red) in blood vessels (green) at P5 in the angiogenic front. Arrowheads point to regions of most intense VEGFR-3 expression in **c** and **d**. **e**, VEGFR-3 (red) and VE-cadherin (green) in a mouse embryo at E9.5. **f**, Intersomitic vessels of a mouse embryo at E9.25. VEGFR-3-positive (red) sprouts are indicated with arrowheads. VE-cadherin staining is in green. DA, dorsal aorta in **f**. Scale bars, 100 µm, except in **e** (which is 500 µm).



Blocking both VEGFR-3 and VEGFR-2 with monoclonal antibodies normalized the vasculature, whereas excess vessels were observed in ears treated with either antibody alone (Fig. 3d, e), indicating that VEGFR-3 signals can sustain a low degree of angiogenesis even in the presence of VEGFR-2 blockers.

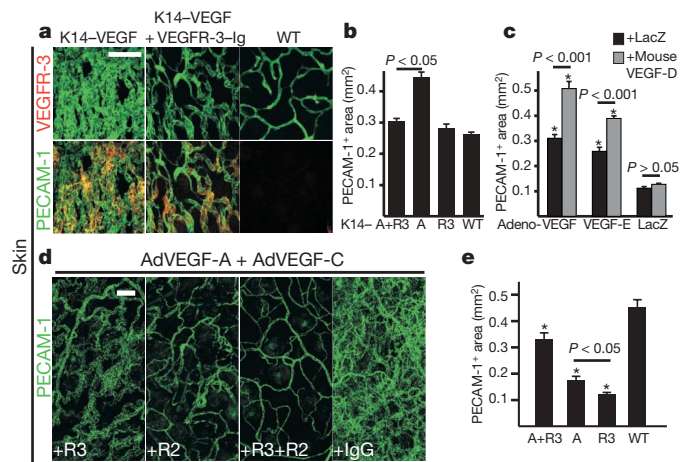
Endothelial Notch activation by DLL4 has recently been shown to suppress excess sprouts in angiogenic vessels<sup>13–16</sup>. Conversely, disruption of a downstream Notch signalling component in zebrafish



resulted in increased sprouting and induced *Vegfr3* (*Flt4*) expression<sup>15</sup>. To study whether Notch regulates VEGFR-3 expression, we treated newborn pups with DAPT, a  $\gamma$ -secretase inhibitor that blocks the Notch signalling pathway. Conversely, we activated the Notch pathway with a small peptide mimetic of the Notch ligand jagged 1 (JAG1; ref. 13). VEGFR-3 expression was not restricted to the angiogenic front in retinas of mice treated with DAPT for 12–48 h (Fig. 4a, b and Supplementary Fig. 10a–c), whereas mice that received JAG1 displayed diminished vascular VEGFR-3 expression (Fig. 4a, b and Supplementary Fig. 10c–e). Furthermore, DAPT induced and JAG1 suppressed *Vegfr3* transcription in the retinas (Fig. 4c), suggesting that the Notch pathway downregulates VEGFR-3 expression *in vivo*, although *VEGFR3* may initially be induced by Notch activation at least in cultured cells<sup>27</sup>. In our experiments, *Vegfr3* mRNA levels did not change until the 12 h time point, whereas *Pdgfr* and *Nrarp*, used as controls for Notch activity, already responded at 6 h (Fig. 4c). This suggests that Notch downregulates *Vegfr3* by means of intermediate effectors such as the transcriptional repressors HEY1, HEY2 or HES1 (ref. 17), which were downregulated at 12 h, when *Vegfr3* upregulation in response to DAPT was first observed (Supplementary Fig. 10f). DAPT also induced VEGFR-3 in tumour blood vessels (Supplementary Fig. 10g), and administration of either VEGFR-3- or VEGFR-2-blocking antibodies suppressed vascular hyperplasia in tumours during DAPT treatment. Additive inhibition of angiogenesis was not seen when combining either receptor-blocking antibody and DAPT (Fig. 4d and data not shown).

Our data indicate that VEGFR-2 activation induces VEGFR-3 (Fig. 3), whereas previous reports have demonstrated that VEGF induces DLL4 in tip cells by means of VEGFR-2 (refs 13, 14 and 16). We found increased VEGFR-3 protein and mRNA levels in the retinas of DLL4 heterozygous pups that survive but exhibit decreased Notch signalling in endothelial cells (Fig. 4e–g), indicating that DLL4 downregulates VEGFR-3 by means of Notch. Intracocularly injected VEGFR-3 antibodies partially rescued the hyperactive sprouting

resulted in increased sprouting and induced *Vegfr3* (*Flt4*) expression<sup>15</sup>. To study whether Notch regulates VEGFR-3 expression, we treated newborn pups with DAPT, a  $\gamma$ -secretase inhibitor that blocks the Notch signalling pathway. Conversely, we activated the Notch pathway with a small peptide mimetic of the Notch ligand jagged 1 (JAG1; ref. 13). VEGFR-3 expression was not restricted to the angiogenic front in retinas of mice treated with DAPT for 12–48 h (Fig. 4a, b and Supplementary Fig. 10a–c), whereas mice that received JAG1 displayed diminished vascular VEGFR-3 expression (Fig. 4a, b and Supplementary Fig. 10c–e). Furthermore, DAPT induced and JAG1 suppressed *Vegfr3* transcription in the retinas (Fig. 4c), suggesting that the Notch pathway downregulates VEGFR-3 expression *in vivo*, although *VEGFR3* may initially be induced by Notch activation at least in cultured cells<sup>27</sup>. In our experiments, *Vegfr3* mRNA levels did not change until the 12 h time point, whereas *Pdgfr* and *Nrarp*, used as controls for Notch activity, already responded at 6 h (Fig. 4c). This suggests that Notch downregulates *Vegfr3* by means of intermediate effectors such as the transcriptional repressors HEY1, HEY2 or HES1 (ref. 17), which were downregulated at 12 h, when *Vegfr3* upregulation in response to DAPT was first observed (Supplementary Fig. 10f). DAPT also induced VEGFR-3 in tumour blood vessels (Supplementary Fig. 10g), and administration of either VEGFR-3- or VEGFR-2-blocking antibodies suppressed vascular hyperplasia in tumours during DAPT treatment. Additive inhibition of angiogenesis was not seen when combining either receptor-blocking antibody and DAPT (Fig. 4d and data not shown).



resulted in increased sprouting and induced *Vegfr3* (*Flt4*) expression<sup>15</sup>. To study whether Notch regulates VEGFR-3 expression, we treated newborn pups with DAPT, a  $\gamma$ -secretase inhibitor that blocks the Notch signalling pathway. Conversely, we activated the Notch pathway with a small peptide mimetic of the Notch ligand jagged 1 (JAG1; ref. 13). VEGFR-3 expression was not restricted to the angiogenic front in retinas of mice treated with DAPT for 12–48 h (Fig. 4a, b and Supplementary Fig. 10a–c), whereas mice that received JAG1 displayed diminished vascular VEGFR-3 expression (Fig. 4a, b and Supplementary Fig. 10c–e). Furthermore, DAPT induced and JAG1 suppressed *Vegfr3* transcription in the retinas (Fig. 4c), suggesting that the Notch pathway downregulates VEGFR-3 expression *in vivo*, although *VEGFR3* may initially be induced by Notch activation at least in cultured cells<sup>27</sup>. In our experiments, *Vegfr3* mRNA levels did not change until the 12 h time point, whereas *Pdgfr* and *Nrarp*, used as controls for Notch activity, already responded at 6 h (Fig. 4c). This suggests that Notch downregulates *Vegfr3* by means of intermediate effectors such as the transcriptional repressors HEY1, HEY2 or HES1 (ref. 17), which were downregulated at 12 h, when *Vegfr3* upregulation in response to DAPT was first observed (Supplementary Fig. 10f). DAPT also induced VEGFR-3 in tumour blood vessels (Supplementary Fig. 10g), and administration of either VEGFR-3- or VEGFR-2-blocking antibodies suppressed vascular hyperplasia in tumours during DAPT treatment. Additive inhibition of angiogenesis was not seen when combining either receptor-blocking antibody and DAPT (Fig. 4d and data not shown).

phenotype seen in DLL4 heterozygotes, as seen by reduced numbers of filopodia and decreased filopodial length (Supplementary Fig. 11a–c). VEGFR-3 antibodies administered during treatment with DAPT also partially inhibited the hyperactive sprouting/branching phenotype induced by Notch inhibition in the retina (Supplementary Fig. 11d, e). These results suggest that upregulation and activation of VEGFR-3 has a functional role in sprouting angiogenesis, and that VEGFR-3 is an important effector of the vascular phenotype resulting from Notch inhibition.

In recent years, the development of angiogenesis inhibitors has focused on the VEGF/VEGFR-2 system. These inhibitors have shown promising efficacy in the clinic in the treatment of tumours and ocular disease, but tumours in particular may eventually become resistant to VEGF/VEGFR-2 inhibitors<sup>28</sup>. Our results indicate that angiogenic sprouting is impaired without VEGFR-3 signals, and suggest that VEGFR-3 may drive angiogenesis even in conditions of therapeutic targeting of VEGFR-2 (Supplementary Fig. 12). VEGF-C is produced by leukocytes, such as macrophages<sup>29</sup>, whereas robust VEGF-C production has been detected in a variety of tumour cell lines and in human tumours, which correlates with increased propensity for lymphatic metastasis in some human tumours<sup>4,22,30</sup>. In light of our results, a significant biological role for VEGF-C in the tumour setting may be to augment angiogenesis and sustain tumour

growth in addition to actively promoting lymphatic metastasis. Thus, our findings elucidate a previously unknown and important regulatory system for angiogenesis, and suggest a new target for anti-angiogenic therapies.

## METHODS SUMMARY

**Animal procedures.** All animal experiments were approved by the Committee for Animal Experiments in the District of Southern Finland. *Vegfr3*-gene-targeted or wild-type E8.5–E11.5 mouse embryos, *Vegfr3*<sup>+/LacZ</sup>, *Vegfr3*<sup>+/LacZ</sup>, *Dll4*<sup>+/LacZ</sup> or wild-type neonatal pups, or adult K14-transgenic or wild-type mice were used for the experiments. Tumour-bearing adult mice or neonatal pups were intraperitoneally or subcutaneously administered VEGFR-3, VEGFR-2, VEGFR-3 + VEGFR-2, or control monoclonal antibodies, and/or DAPT, JAG1, scrambled control peptide (SC-JAG1) or vehicle. Alternatively, tumour-bearing mice were injected intravenously with adenoviruses encoding VEGFR-3-Ig or  $\beta$ -galactosidase as a control. Ears of adult immunodeficient nu/nu mice were injected with adenoviral vectors encoding VEGF family growth factors or control.

**Tissue imaging.** Embryos, postnatal retinas or skin were processed for whole-mount staining, whereas tumours were cut into 50- $\mu$ m sections for immunostaining. Proliferating or hypoxic cells were identified by injection of bromodeoxyuridine or pimonidazole, respectively, 1–2 h before euthanasia. Commercially available or previously published antibodies were used for immunostaining according to standard protocols. Samples were imaged using a Zeiss Axiovert epifluorescence or a Zeiss LSM 510 Meta/LIVE5 confocal laser-scanning microscope.

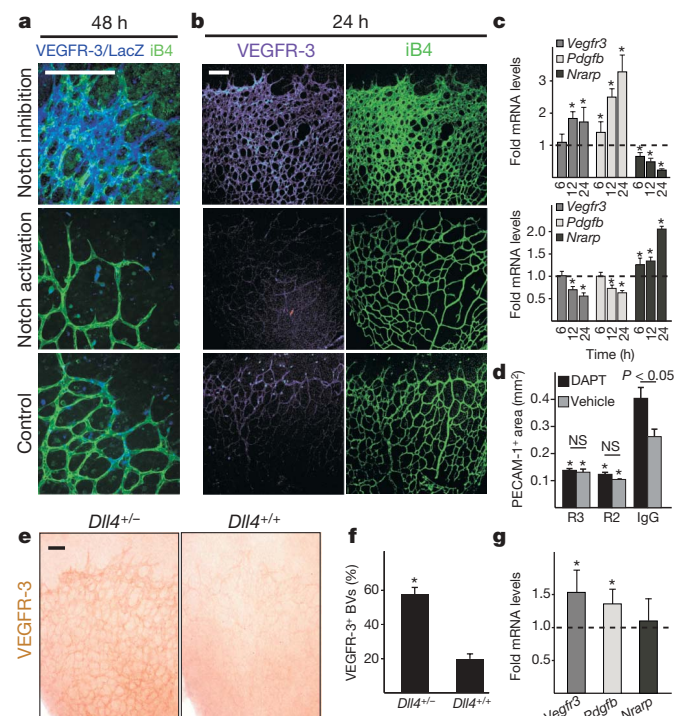
**Molecular analyses.** Gene expression levels were analysed from RNA extracted from postnatal retinas at P5 by quantitative real-time PCR, as described previously<sup>13</sup>.

**Quantitative analysis.** At least three animals were analysed in each time point and treatment. Statistical analysis was carried out using the unpaired Student's *t*-test or one-way analysis of variance (ANOVA).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 2 July 2007; accepted 13 May 2008.

Published online 25 June 2008.



**Figure 4 | Notch signalling downregulates VEGFR-3 in endothelial cells.**

**a–c,** Analysis of the effects of the  $\gamma$ -secretase inhibitor DAPT (Notch inhibition) or the JAG1 peptide (Notch activation) on VEGFR-3 expression in the postnatal mouse retina. VEGFR-3 (blue pseudocolour) in the retinas of *Vegfr3*<sup>+/LacZ</sup> heterozygous mice (**a**), and VEGFR-3 (purple rainbow colour) and isolectin B4 (green) staining (**b**), in wild type retinas at P5 are shown. Quantification of changes in *Vegfr3*, *Pdgfrb* and *Nrarp* mRNA levels after DAPT (upper graph) or JAG1 (lower graph) administration is shown in **c**. The dashed line indicates the relative mRNA levels in vehicle- or SC-JAG1-treated controls ( $n \geq 3$  mice per group, error bars indicate  $\pm$  s.d.). **d,** Quantification of the PECAM-1-positive surface area in tumours treated with DAPT or vehicle and the indicated monoclonal antibodies ( $n \geq 4$  tumours per group; error bars indicate  $\pm$  s.e.m.). **e,** Simultaneous VEGFR-3 immunostaining of *Dll4*<sup>+/LacZ</sup> and wild-type littermate retinas. **f–g,** Quantification of the VEGFR-3-positive surface area (**f**) and changes in *Vegfr3*, *Pdgfrb* and *Nrarp* mRNA levels (**g**) in *Dll4*<sup>+/LacZ</sup> and *Dll4*<sup>+/+</sup> pups ( $n \geq 3$  mice per group). Error bars indicate  $\pm$  s.e.m. in **f** and  $\pm$  s.d. in **g**. NS, not statistically significant; asterisk,  $P < 0.05$ ; scale bars, 100  $\mu$ m.

- Carmeliet, P. Angiogenesis in life, disease and medicine. *Nature* **438**, 932–936 (2005).
- Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* **9**, 669–676 (2003).
- Kaipainen, A. et al. Expression of the fms-like tyrosine kinase 4 gene becomes restricted to lymphatic endothelium during development. *Proc. Natl Acad. Sci. USA* **92**, 3566–3570 (1995).
- Valtola, R. et al. VEGFR-3 and its ligand VEGF-C are associated with angiogenesis in breast cancer. *Am. J. Pathol.* **154**, 1381–1390 (1999).
- Paavonen, K., Puolakkainen, P., Jussila, L., Jähkola, T. & Alitalo, K. Vascular endothelial growth factor receptor-3 in lymphangiogenesis in wound healing. *Am. J. Pathol.* **156**, 1499–1504 (2000).
- Ferrara, N. et al. Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* **380**, 439–442 (1996).
- Carmeliet, P. et al. Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* **380**, 435–439 (1996).
- Shalaby, F. et al. Failure of blood island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–66 (1995).
- Gille, H. et al. Analysis of biological effects and signaling properties of Flt-1 (VEGFR-1) and KDR (VEGFR-2). A reassessment using novel receptor-specific vascular endothelial growth factor mutants. *J. Biol. Chem.* **276**, 3222–3230 (2001).
- Gerhardt, H. et al. VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163–1177 (2003).
- Noguera-Troise, I. et al. Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis. *Nature* **444**, 1032–1037 (2006).
- Ridgway, J. et al. Inhibition of Dll4 signalling inhibits tumour growth by deregulating angiogenesis. *Nature* **444**, 1083–1087 (2006).
- Hellstrom, M. et al. Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature* **445**, 776–780 (2007).
- Lobov, I. B. et al. Delta-like ligand 4 (Dll4) is induced by VEGF as a negative regulator of angiogenic sprouting. *Proc. Natl Acad. Sci. USA* **104**, 3219–3224 (2007).
- Siekmann, A. F. & Lawson, N. D. Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries. *Nature* **445**, 781–784 (2007).
- Suchting, S. et al. The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proc. Natl Acad. Sci. USA* **104**, 3225–3230 (2007).
- Roca, C. & Adams, R. H. Regulation of vascular morphogenesis by Notch signaling. *Genes Dev.* **21**, 2511–2524 (2007).

18. Alitalo, K., Tammela, T. & Petrova, T. V. Lymphangiogenesis in development and human disease. *Nature* **438**, 946–953 (2005).
19. Dixelius, J. *et al.* Ligand-induced vascular endothelial growth factor receptor-3 (VEGFR-3) heterodimerization with VEGFR-2 in primary lymphatic endothelial cells regulates tyrosine phosphorylation sites. *J. Biol. Chem.* **278**, 40973–40979 (2003).
20. Dumont, D. J. *et al.* Cardiovascular failure in mouse embryos deficient in VEGF receptor-3. *Science* **282**, 946–949 (1998).
21. Covassin, L. D., Villefranc, J. A., Kacergis, M. C., Weinstein, B. M. & Lawson, N. D. Distinct genetic interactions between multiple Vegf receptors are required for development of different blood vessel types in zebrafish. *Proc. Natl Acad. Sci. USA* **103**, 6554–6559 (2006).
22. Laakkonen, P. *et al.* Vascular endothelial growth factor receptor 3 is involved in tumor angiogenesis and growth. *Cancer Res.* **67**, 593–599 (2007).
23. Veikkola, T. *et al.* Intrinsic versus microenvironmental regulation of lymphatic endothelial cell phenotype and function. *FASEB J.* **17**, 2006–2013 (2003).
24. Goldman, J. *et al.* Cooperative and redundant roles of VEGFR-2 and VEGFR-3 signaling in adult lymphangiogenesis. *FASEB J.* **21**, 1003–1012 (2007).
25. Makinen, T. *et al.* Inhibition of lymphangiogenesis with resulting lymphedema in transgenic mice expressing soluble VEGF receptor-3. *Nature Med.* **7**, 199–205 (2001).
26. Baldwin, M. E. *et al.* The specificity of receptor binding by vascular endothelial growth factor-D is different in mouse and man. *J. Biol. Chem.* **276**, 19166–19171 (2001).
27. Shawber, C. J. *et al.* Notch alters VEGF responsiveness in human and murine endothelial cells by direct regulation of VEGFR-3 expression. *J. Clin. Invest.* **117**, 3369–3382 (2007).
28. Ferrara, N. & Kerbel, R. S. Angiogenesis as a therapeutic target. *Nature* **438**, 967–974 (2005).
29. Baluk, P. *et al.* Pathogenesis of persistent lymphatic vessel hyperplasia in chronic airway inflammation. *J. Clin. Invest.* **115**, 247–257 (2005).
30. Stacker, S. A., Achen, M. G., Jussila, L., Baldwin, M. E. & Alitalo, K. Lymphangiogenesis and cancer metastasis. *Nature Rev. Cancer* **2**, 573–583 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to thank P. Haiko, W. Holnthoner, T. Holopainen and D. Tvorogov for help with the experiments, as well as A. Ristimäki for the MKN45 cells, and T. Takahashi for NCI-H460-LNM35 cells. We also thank S. Fanta, C. Heckman, K. Helenius and T. Petrova for critical

comments on the manuscript. The Biomedicum Molecular Imaging Unit is acknowledged for microscopy services, and M. Helanterä, P. Hyvärinen, A. Kotronen, T. Laakkonen, S. Lampi, K. Makkonen, A. Malinen, T. Tainola and S. Wallin for technical assistance, as well as A. Lehtonen and T. Taina for animal husbandry. Electron microscopy was carried out in collaboration with the Electron Microscopy Unit, Institute of Biotechnology at the University of Helsinki. This work was supported by grants from the NIH (5 R01 HL075183-02), The European Union (Lymphangiogenomics, LSHG-CT-2004-503573) and the Louis Jeantet Foundation (K.A.), as well as the Association for International Cancer Research (UK) and IngaBritt and Anne Lundberg Foundation (C.B.). T.T. was supported by personal grants from the Finnish Cancer Organizations, the Finnish Cultural Foundation, Nylands Nation, The Paulo Foundation and the Helsinki Biomedical Graduate School.

**Author Contributions** T.T. designed, directed and performed embryo, retina, mouse ear and xenograft experiments, immunohistochemistry and data analysis, interpreted results and wrote the paper; G.Z. performed xenograft experiments, immunohistochemistry, electron microscopy and ELISA, analysed data and interpreted results; E.W. designed and performed qRT-PCR and data analysis, and interpreted results; A.M. performed embryo and retina experiments, immunohistochemistry, and data analysis; S.S. performed intraocular injections and immunohistochemistry, analysed data and interpreted results; M. Wirzenius designed and performed mouse ear experiments and immunohistochemistry, analysed data and interpreted results; M. Waltari performed xenograft experiments, immunohistochemistry and data analysis; M. H. directed experiments and interpreted results; T.S. performed Rip1Tag2 tumour experiments, analysed data and interpreted results; R.P. performed immunohistochemistry and analysed data; C.F. performed intraocular injections; A.D. provided the *Dll4*<sup>+/-</sup> mice; H.I. performed surgery and provided clinical tumour samples; P.L. directed experiments and interpreted results; G.C. directed experiments and interpreted results; S.Y.-H. developed and provided adenovirus vectors; M.S. generated and provided K14-VEGF and K14-VEGF-E transgenic mice; B.P. generated and provided monoclonal VEGFR function-blocking antibodies; A.E. directed experiments and interpreted results; C.B. designed experiments, interpreted results and helped write the paper; K.A. designed experiments, interpreted results and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the paper on [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.A. ([kari.alitalo@helsinki.fi](mailto:kari.alitalo@helsinki.fi)).



## METHODS

**Mice and tissues.** The study was approved by the Committee for Animal Experiments of the District of Southern Finland. For tumour injections and before they were killed, the mice were anesthetized with intraperitoneal injections of xylazine (10 mg kg<sup>-1</sup>) and ketamine (50 mg kg<sup>-1</sup>). The *Vegfr3<sup>+/-LacZ</sup>* (ref. 20), *Vegfr3<sup>+/-LacZ</sup>* (ref. 31) and *Dll4<sup>+/-LacZ</sup>* (ref. 32) knock-in mice, as well as the K14-VEGF-E (ref. 33), K14-VEGF<sub>165</sub> (ref. 34), K14-VEGFR-3-Ig (ref. 25) and Rip1Tag2 (refs 35 and 36) transgenic mouse lines, were as published previously. T-cell-deficient Balb/c nu/nu or NMRI nu/nu mice (Taconic) were used for the tumour xenografting experiments. Neonatal wild-type mice in the NMRI, CD1 or C57black background, *Vegfr3<sup>+/-LacZ</sup>* mice in the NMRI background, *Vegfr3<sup>+/-LacZ</sup>* and *Vegfr3<sup>+/-LacZ</sup>* mice in the ICR background, or *Dll4<sup>+/-LacZ</sup>* mice in the CD1 background were used for the experiments. Twenty-four-day-old wild-type female FVB/n mice were used for superovulation experiments. After killing the mice, tissues were immersed in 4% paraformaldehyde, washed in phosphate buffered saline (PBS) and then processed for whole-mount staining, immersed in OCT medium (Tissue Tek) or embedded in paraffin.

**Analysis of postnatal angiogenesis in the mouse retina.** Neonatal wild-type mice in the NMRI background were intraperitoneally or subcutaneously injected with 50 mg kg<sup>-1</sup> d<sup>-1</sup> of anti-VEGFR-3 (mF4-31C1, ref. 37) or anti-VEGFR-2 (DC101, ref. 38) monoclonal antibodies, which inhibit ligand binding to mouse VEGFR-3 or VEGFR-2, respectively. Non-specific rat IgG (Dako) was used as a control. Alternatively, 50 mg kg<sup>-1</sup> d<sup>-1</sup> anti-VEGFR-3 plus 50 mg kg<sup>-1</sup> d<sup>-1</sup> anti-VEGFR-2 were given in combination. Treatment with the antibodies was continued from P0 until P5, when the mice were killed and their eyes collected for analysis. Intraocular injections were carried out as described previously<sup>16</sup>, and the pups were killed 5 h after injection. The  $\gamma$ -secretase inhibitor DAPT (*N*-[*N*-(3,5-difluorophenacetyl-L-alanyl)]-S-phenylglycine *t*-butyl ester, Sigma)<sup>39</sup> was dissolved in 10% ethanol and 90% sunflower oil (Sigma), and injected subcutaneously at 100 mg kg<sup>-1</sup> in a volume of 10  $\mu$ l g<sup>-1</sup> into P2 pups twice daily. The vehicle was used as a negative control. The small peptide mimetic of the Notch ligand JAG1 or SC-JAG1 (Thermo Scientific) was dissolved in 50% dimethylsulphoxide (DMSO) and 50% sterile water, and was administered subcutaneously at 10 mg kg<sup>-1</sup> (refs 13, 40). Pups were killed on P5, 6 h, 12 h, 24 h or 48 h after administration of DAPT, JAG1, SC-JAG1 or vehicle at 12-h intervals. Alternatively, the mice were given DAPT four times and the monoclonal antibodies twice over a period of 48 h. For labelling of proliferating cells, the mice were given an intraperitoneal injection of bromodeoxyuridine (1 mg kg<sup>-1</sup>, Sigma) 2 h before they were killed.

**Transduction of the mouse ear skin with adenoviral gene transfer vectors.** Adenoviruses encoding human VEGF<sub>165</sub> (ref. 41), human full-length VEGF-C (ref. 41), mouse VEGF-D or the Orf(NZ7) virus-encoded VEGF-E (ref. 42) were injected intradermally into the ears of immunodeficient nu/nu mice in the following combinations: VEGF + LacZ, VEGF + mVEGF-D, VEGF-E + LacZ, VEGF-E + mVEGF-D or mVEGF-D + LacZ, or LacZ alone. AdLacZ (ref. 41) was used as a negative control. 10<sup>8</sup> plaque-forming units (p.f.u.) of each virus (altogether 2  $\times$  10<sup>8</sup> p.f.u.) were injected in a volume of 50  $\mu$ l. The ears were harvested 6 days after injection and processed for whole-mount analysis, as described previously<sup>43</sup>. Alternatively, 10<sup>8</sup> p.f.u. of both AdVEGF and AdVEGF-C were injected on day 0. The mice were given anti-VEGFR-3 (40 mg kg<sup>-1</sup>), anti-VEGFR-2 (40 mg kg<sup>-1</sup>), anti-VEGFR-3 and anti-VEGFR-2 (40 + 40 mg kg<sup>-1</sup>) in combination, or IgG (40 mg kg<sup>-1</sup>) on days 4 and 6. The mice were killed on day 8 and ears were collected for analysis.

**Human samples.** Patients diagnosed with disseminated adenocarcinoma of the colon underwent a standard chemotherapy regimen, followed by partial resection of the liver to eradicate metastases. Samples from the interface between normal and tumour tissue were immediately immersed in 4% paraformaldehyde overnight (12–16 h) at +4 °C, washed in PBS, incubated in 25% sucrose overnight at +4 °C, embedded in optimal cutting temperature (OCT) medium and frozen. 50- $\mu$ m sections were used for staining. The study was approved by the Ethical Committee of The Hospital District of Helsinki and Uusimaa according to the guidelines of the Helsinki declaration. Informed consent was obtained from the patients.

**Tumour cell lines, xenografts and treatments.** NCI-H460-LNM35 cells (LNM35), a sub-line of NCI-H460-N15, a human large-cell carcinoma of the lung<sup>44</sup> (a gift from T. Takahashi), MKN45 human gastric carcinoma cells<sup>45</sup> (a gift from A. Ristimäki), B16 mouse melanoma or Lewis mouse lung carcinoma cells were maintained in RPMI-1640 medium, whereas the G401 kidney cancer cells (ATCC CRL-1441) were grown in DMEM. Both media were supplemented with 2 mM L-glutamine, penicillin (100 U ml<sup>-1</sup>), streptomycin (100  $\mu$ g ml<sup>-1</sup>) and 10% fetal bovine serum (Autogen Bioclear).

The LNM35, MKN45 and G401 xenografts, as well as the B16 and LLC syngeneic grafts, were made by injecting 0.5–5  $\times$  10<sup>6</sup> cells into the subcutaneous

space in the abdominal flank of immunodeficient mice. For experiments involving antibody treatment, 10<sup>6</sup> LNM35 cells or 5  $\times$  10<sup>5</sup> B16 cells were injected into the subcutaneous space above the abdomen of each mouse. Tumours were allowed to reach a volume of approximately 50 mm<sup>3</sup> and then the mice were randomized by tumour volume into treatment groups. The tumour-bearing mice were treated with anti-VEGFR-3, anti-VEGFR-2 or non-specific rat IgG (Dako) every other day by intraperitoneal injection (40 mg kg<sup>-1</sup>). Mice that received anti-VEGFR-3 and anti-VEGFR-2 in combination were given 40 mg kg<sup>-1</sup> of each antibody. Alternatively, at the time of tumour implantation, mice were intravenously administered adenoviral soluble VEGFR-3-Ig (ref. 46), or control adenovirus encoding  $\beta$ -galactosidase (1  $\times$  10<sup>9</sup> p.f.u. per mouse). For analysis of the short-term effects on the vasculature and stroma of LNM35 or B16 tumours, tumour-bearing mice were treated for 6 days with AdVEGFR-3-Ig, antibodies and/or DAPT (100 mg kg<sup>-1</sup> d<sup>-1</sup>). Long-term effects were evaluated in LNM35 tumour-bearing mice by administering antibodies and/or DAPT (75 mg kg<sup>-1</sup> d<sup>-1</sup>) until the tumours reached a maximal size of 20 mm in diameter. Tumours were measured in x, y and z dimensions using a digital caliper, and volumes were calculated using the ellipse formula ( $V = xyz/2$ ). Mice were given 60 mg kg<sup>-1</sup> pimonidazole (Chemicon) intraperitoneally 1 h before they were killed.

**Treatment of Rip1Tag2 transgenic mice with monoclonal antibodies.** Altogether 12 Rip1Tag2 transgenic mice<sup>35,36</sup> were treated with 30 mg kg<sup>-1</sup> anti-VEGFR-3, whereas 6 mice received anti-VEGFR-2 and 15 mice the control IgG every other day for two weeks. Mice were killed on day 14. For the analysis of blood vessel density and blood vessel functionality, mice were tail-vein-injected with 100  $\mu$ l of 1 mg ml<sup>-1</sup> fluorescein-labelled *Lycopersicon esculentum* lectin (Vector Laboratories) under inhalation-anaesthesia with isoflurane (Minrad Inc.). After 5 min, mice were perfused with 10 ml of 4% paraformaldehyde followed by 10 ml PBS. Isolated pancreata were immersed in ascending concentrations of sucrose (12%, 15% and 18%, for 1 h each), embedded in OCT (Tissue-Tek) and snap-frozen in liquid nitrogen. For the evaluation of blood vessel density, 10- $\mu$ m sections were analysed with a Nikon Diaphot 300 immunofluorescence microscope using Openlab 3.1.7. Software (Improvision) and the number of fluorescein-labelled vessels in the tumours was determined using Image J software (The National Institutes of Health).

**Superovulation.** 24-day-old mice received an intraperitoneal injection of 5 international units (IU) pregnant mare serum gonadotropin (National Hormone and Peptide Program, Harbor-UCLA Medical Centre) followed by an injection of 5 IU human chorionic gonadotropin (Pregnyl) 47 h later. Mice were killed and ovaries collected for analysis 72 or 84 h after stimulation of superovulation.

**Immunohistochemistry.** 50- $\mu$ m sections of tumours or ovaries, 80- $\mu$ m cross-sections of the duodenum, and 7- $\mu$ m sections of skin were fixed with cold acetone, washed with PBS and blocked with TNB (PerkinElmer). The following primary antibodies were used for immunostaining of mouse tissues: polyclonal goat anti-mouse VEGFR-3 (AF743, R&D Systems), rat-anti mouse VEGFR-3 (clone AFL4, 16-5988, eBioscience), polyclonal goat anti-mouse VEGFR-2 (AF644, R&D Systems), rat anti-mouse VEGFR-2 (clone AVAS-12 $\alpha$ 1, 550549, BD Biosciences), rat anti-mouse VEGFR-1 (clone 5B12, ImClone), polyclonal goat anti-mouse DLL4 (AF1389, R&D Systems), hamster anti-PECAM-1 (clone 2H8, MAB1398Z, Chemicon), unconjugated or FITC-conjugated rat anti-PECAM-1 (clone MEC 13.3, 557355 and 553372, BD Biosciences), FITC-conjugated CD11b (clone M1/70, 550282, BD Biosciences), polyclonal rabbit anti-FITC (71-1900, Zymed/Invitrogen), polyclonal rabbit anti-VEGF-C (6 and pre-immune serum<sup>29</sup>), rat anti-mouse VE-cadherin (clone 11D4.1, 550548, BD Biosciences) rat anti-mouse pan-endothelial antigen (clone MECA-32, 550563, BD Biosciences), polyclonal rabbit anti-cow GFAP (Z0334, Dako), rat anti-mouse F4/80 (clone BM8, BM4007, Acris antibodies), polyclonal rabbit anti-NG2 (AB5320, Chemicon), Cy3-conjugated mouse anti-SMA (clone 1A4, C6189, Sigma), polyclonal rabbit anti-LYVE-1 (ref. 31) and FITC-conjugated mouse anti-pimonidazole (Hypoxypore, HP2-1000, Chemicon). Human tissues were stained with mouse anti-human VEGFR-3 (clone 9D9, ref. 47), polyclonal rabbit anti-podoplanin<sup>48</sup>, or polyclonal rabbit anti-human von Willebrand factor (A082, Dako). Sections were washed with TNT-buffer (0.15 M NaCl, 0.1 M Tris-HCl, pH 7.5, 0.05% Tween20) and the primary antibodies were detected with the appropriate Alexa 488, 594 or 647 secondary antibody conjugates (Molecular Probes/Invitrogen). Bromodeoxyuridine (BrdU) was detected with Alexa 594-conjugated mouse anti-BrdU antibodies (Molecular Probes/Invitrogen) after incubation in hydrochloric acid and neutralization using sodium tetraborate. For whole-mount analysis of the cutaneous blood vessels, dermal tissues of the ear<sup>43</sup> or ventral skin of pups<sup>49</sup> were exposed for staining, as published. For analysis of the microvasculature, retinas were stained with biotinylated *Griffonia simplicifolia* lectin (isolectin B4, Vector Laboratories), as before<sup>10</sup>, followed by immunostaining. Alternatively, retinas,

X-gal-stained *Vegf*<sup>+/LacZ</sup> embryos or tumour sections were incubated in 5% hydrogen peroxide in methanol, followed by VEGFR-3 or PECAM-1 antibodies, biotinylated secondary antibodies (Vector Laboratories) and signal detection with the Avidin-Biotin Complex kit (Vector Laboratories)<sup>50</sup> or tyramide signal amplification (Perkin Elmer). Diaminobenzidine (Sigma or Chemicon) was used as the chromogen in both protocols. X-gal- and PECAM-1-stained *Vegf*<sup>+/LacZ</sup> embryos were embedded in paraffin and sectioned in horizontal or sagittal orientation. Nuclear counterstaining was carried out with nuclear red<sup>31</sup>.

The specificity of the goat anti-mouse VEGFR-3 antibody (R&D Systems) to VEGFR-3 was verified by whole-mount staining of VEGFR-3-knockout embryos collected at E9, and compared with wild-type embryos (Supplementary Fig. 6c and data not shown). We observed specific staining in the vessels in wild-type embryos, whereas no staining could be observed in VEGFR-3-deficient embryos, indicating that the VEGFR-3 antibody used in the study was indeed specific for VEGFR-3.

All fluorescently labelled samples were mounted with Vectashield mounting medium containing 4,6-diamidino-2-phenylindole (DAPI; H-1200, Vector Laboratories), and analysed with a compound fluorescent microscope (Zeiss 2, Carl Zeiss;  $\times 10$  objective with numerical aperture 0.30) or a confocal microscope (Zeiss LSM 510, oil objectives  $\times 40$  with NA 1.3 and  $\times 63$  with NA 1.4) by using multichannel scanning in frame mode. The pinhole diameter was set at 1 Airy unit for detection of the Alexa 488 signal, and was adjusted for identical optical slice thickness for the fluorochromes emitting at higher wavelengths. Three-dimensional projections were digitally reconstructed from confocal z-stacks. Co-localization of signals was assessed from single confocal optical sections. Images of whole embryos were acquired using tile scanning using a pinhole diameter  $>3.0$  Airy units. X-gal staining of LacZ reporter mice was carried out as published<sup>31</sup>, and samples were mounted in Aquamount (BDH Laboratory Supplies) or Vectashield mounting medium containing DAPI. Co-localization of the  $\beta$ -galactosidase chromogen and the fluorescent signal was carried out as described previously<sup>13</sup>.

**Analysis of kidney pathology and function.** NMRI mice received anti-VEGFR-3, anti-VEGFR-2 or non-specific rat IgG (40 mg kg<sup>-1</sup> every other day), or both anti-VEGFR-3 and anti-VEGFR-2 (40 mg kg<sup>-1</sup> of each antibody every other day), for 10 days ( $n \geq 3$ ). Mice were killed and urine was collected by bladder puncture using a 33-gauge needle. Proteinuria was assessed by pipetting 10  $\mu$ l of urine on colour-coded strips (Combur 3 Test Strips, Roche). We did not observe significant proteinuria in the mice by this method (data not shown). Albuminuria was measured using an indirect ELISA kit for detection of mouse albumin (Albuwell M, Exocell) according to the manufacturer's instructions. The albumin levels were normalized to urine creatinine levels using a direct ELISA assay (R&D Systems). Kidneys were collected, cut in half in the horizontal plane at the pyloric midline, fixed and embedded in paraffin. Paraffin sections were stained with haematoxylin/eosin or PECAM-1 antibodies, and observed under a bright field microscope. For transmission electron microscopy (TEM), at least three antibody-treated NMRI mice or K14-VEGFR-3-Ig transgenic mice were perfusion-fixed with 4% formaldehyde in 100 mM phosphate buffer, pH 7.4, and slices of kidney were incubated in the same fixative overnight (16 h) at +10 °C. Samples were then fixed with 2% glutaraldehyde in the same buffer for 1 h, and post-fixed with 1% buffered osmium tetroxide for 1 h, dehydrated and embedded in epon at room temperature (+22 °C). Sections were post-stained with uranyl acetate and lead and were examined with a Jeol EX1200 II TEM operating at 60 kV. Images were acquired with an ES500W CCD camera (Gatan Corp.).

**Vessel morphometry and quantitative analysis.** The vascular surface area in retinas was quantified as an isolectin B4-positive area from  $\times 10$  confocal micrographs of all intact quarters of the processed retina using Image J software, as described previously<sup>42,43</sup>. PECAM-1-positive vessels in the ears or the LYVE-1-positive vessel area in the intestines were quantified in a similar manner. VEGFR-3-positive surface area was normalized to isolectin B4 surface area from double-stained retinas to yield the percentage of VEGFR-3-positive vessels. Vessel branching points and sprouts were counted manually as in ref. 14, and their number was normalized to the area of retina analysed. This analysis was carried out from 22 anti-VEGFR-3-, 18 anti-VEGFR-2-, 24 IgG- and 12 anti-VEGFR-3 plus anti-VEGFR-2-treated retinas from four separate experiments. BrdU-positive endothelial cells were counted manually from  $\times 10$  confocal z-stacks of BrdU and isolectin B4 double-stained retinas. Again, all intact quarters of the retina were used for analysis. Endothelial sprout number and length was counted from fluorescent micrographs of *Dll4*<sup>+/-</sup> heterozygous retinas, as described previously<sup>16</sup>. Four *Dll4*<sup>+/-</sup> mice received anti-VEGFR-3, three received anti-VEGFR-2 and two received control IgG.

The embryos were staged at E8.5–E9.5 according to the Edinburgh Mouse Atlas criteria. Six litters at E8.5 ( $n = 6$  -/-, 6 +/+), six litters at E9.0 ( $n = 8$  -/-, 10 +/+) and three litters at E9.5 ( $n = 5$  -/-, 5 +/+) were analysed. The

number of vessel branching points was manually counted from  $\times 10$  confocal stacks in the head at E8.5, where one layer of vessels can be seen (Supplementary Fig. 6a). Again, the criteria used in ref. 14 were applied. The length and area of intersomitic vessels from between somites 12 and 15 was quantified from  $\times 10$  micrographs obtained with the pinhole open.

Tumour vasculature was analysed from thick sections stained with PECAM-1 antibodies. At least three 0.81 mm<sup>2</sup> or 1.69 mm<sup>2</sup> micrographs from regions of uniform staining intensity were acquired, and vascular surface area was quantified using Image J software. According to our analysis, all vessel sprouts in LNM35 tumours were VEGFR-3-positive, whereas most sprouts expressed low levels of the junctional protein PECAM-1. As a result, we used VEGFR-3 as a marker for sprouts in tumours. The sprouts were counted manually from LNM35 tumour sections from regions in parts of the tumour that did not contain LYVE-1-positive lymphatic vessels<sup>22</sup>. VEGFR-3 did not co-localize with pericyte markers (smooth muscle actin, PDGFR- $\beta$ , NG2), nor was it expressed by any of the tumour cells analysed<sup>22</sup>, or by Rip1Tag2 insulinoma cells (data not shown). Approximately 5% of cells expressing the macrophage/monocyte marker CD11b were also VEGFR-3-positive. Vessels positive for lectin were counted from Rip1Tag2 tumour sections. Images were edited using PhotoShop software (Adobe).

**Real-time quantitative PCR.** RNA was isolated from retinas using the RNeasy Micro Kit (Qiagen, 74004). Homogenization was performed using rotor-stator homogenization, followed by disruption using a shredder column (Qiagen, 79656) and on-column DNase digestion. Samples were quality controlled on an Agilent 2100 Bioanalyser using the Agilent 2100 Expert software (version B.02.02.SI238) and the Agilent RNA 6000 Nano Kit (5067-1511). Two separate *in vitro* transcription reactions were performed from every RNA sample using the SuperScript III First-Strand Synthesis Kit for real-time (RT)-PCR (Invitrogen, 18080-051) according to the manufacturer's instructions. Three RT quantitative PCR reactions/TaqMan assays were carried out from every *in vitro* transcription reaction using TaqMan Universal PCR Master Mix (43004437, Applied Biosystems) and the 7300 Real Time PCR System (Applied Biosystems) according to a standardized protocol<sup>13</sup>. The TaqMan Assays (Applied Biosystems) used: *Gapdh* (4352932E), *Cdh5* (Mm00486938\_m1), *Pdgfrb* (Mm00440678\_m1), *Vegfr2* (Mm001222419\_m1), *Vegfr3* (Mm00433337\_m1), *Nrarp* (Mm00482529\_s1), *Vegfa* (Mm00437304\_m1), *Hey1* (Mm00468865\_m1), *Hey2* (Mm00469280\_m1) and *Hes1* (Mm00456108\_g1). Using the Sequence Detection System (version 1.4, Applied Biosystems), the data were normalized to the endogenous control *Gapdh*, and fold changes were calculated using the comparative CT method. In the case of *Vegfr3* and *Pdgfrb* in DAPT-treated retinas, relative fold changes in *Cdh5* were used to control for increased endothelial cell numbers in the sample (Supplementary Fig. 10h). At least three retinas from pups treated with DAPT, JAG1, SC-JAG1 or DAPT vehicle (10% ethanol, 90% sunflower oil) from each time point (6 h, 12 h, 24 h and 48 h) were used for analysis at P5. The contralateral retina was analysed by VEGFR-3 immunofluorescence and isolectin B4 staining. Altogether five retinas from *Dll4* heterozygous (*Dll4*<sup>+/-</sup>) and five retinas from wild-type littermate pups (*Dll4*<sup>+/+</sup>) were used for the analysis.

**Statistical analysis.** At least six animals were used for each analytical time point and technique studied. Representative data from at least three replicate experiments are shown, with the exception of the experiments involving *Dll4*<sup>+/-</sup> mice and the Rip1Tag2 transgenic mice. A two-tailed Student's *t*-test or two-way ANOVA was used for statistical analysis. A *P* value of less than 0.05 was considered to be statistically significant, and is indicated with an asterisk in the figures.

- Karkkainen, M. J. *et al.* Vascular endothelial growth factor C is required for sprouting of the first lymphatic vessels from embryonic veins. *Nature Immunol.* **5**, 74–80 (2004).
- Duarte, A. *et al.* Dosage-sensitive requirement for mouse *Dll4* in artery development. *Genes Dev.* **18**, 2474–2478 (2004).
- Kiba, A., Sagara, H., Hara, T. & Shibuya, M. VEGFR-2-specific ligand VEGF-E induces non-edematous hyper-vascularization in mice. *Biochem. Biophys. Res. Commun.* **301**, 371–377 (2003).
- Zheng, Y. *et al.* Chimeric VEGF-ENZ7/PIGF promotes angiogenesis via VEGFR-2 without significant enhancement of vascular permeability and inflammation. *Arterioscler. Thromb. Vasc. Biol.* **26**, 2019–2026 (2006).
- Hanahan, D. Heritable formation of pancreatic beta-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. *Nature* **315**, 115–122 (1985).
- Esni, F. *et al.* Neural cell adhesion molecule (N-CAM) is required for cell type segregation and normal ultrastructure in pancreatic islets. *J. Cell Biol.* **144**, 325–337 (1999).
- Pytowski, B. *et al.* Complete and specific inhibition of adult lymphatic regeneration by a novel VEGFR-3 neutralizing antibody. *J. Natl Cancer Inst.* **97**, 14–21 (2005).

38. Prewett, M. *et al.* Antivascular endothelial growth factor receptor (fetal liver kinase 1) monoclonal antibody inhibits tumor angiogenesis and growth of several mouse and human tumors. *Cancer Res.* **59**, 5209–5218 (1999).
39. Dovey, H. F. *et al.* Functional gamma-secretase inhibitors reduce  $\beta$ -amyloid peptide levels in brain. *J. Neurochem.* **76**, 173–181 (2001).
40. Weijzen, S. *et al.* The Notch ligand Jagged-1 is able to induce maturation of monocyte-derived human dendritic cells. *J. Immunol.* **169**, 4273–4278 (2002).
41. Enholm, B. *et al.* Adenoviral expression of vascular endothelial growth factor-C induces lymphangiogenesis in the skin. *Circ. Res.* **88**, 623–629 (2001).
42. Wirzenius, M. *et al.* Distinct vascular endothelial growth factor signals for lymphatic vessel enlargement and sprouting. *J. Exp. Med.* **204**, 1431–1440 (2007).
43. Tammela, T. *et al.* Angiopoietin-1 promotes lymphatic sprouting and hyperplasia. *Blood* **105**, 4642–4648 (2005).
44. Kozaki, K. *et al.* Establishment and characterization of a human lung cancer cell line NCI-H460-LNM35 with consistent lymphogenous metastasis via both subcutaneous and orthotopic propagation. *Cancer Res.* **60**, 2535–2540 (2000).
45. Sakai, K. *et al.* Expression and function of class II antigens on gastric carcinoma cells and gastric epithelia: differential expression of DR, DQ, and DP antigens. *J. Natl Cancer Inst.* **79**, 923–932 (1987).
46. Karpanen, T. *et al.* Vascular endothelial growth factor C promotes tumor lymphangiogenesis and intralymphatic tumor growth. *Cancer Res.* **61**, 1786–1790 (2001).
47. Jussila, L. *et al.* Lymphatic endothelium and Kaposi's sarcoma spindle cells detected by antibodies against the vascular endothelial growth factor receptor-3. *Cancer Res.* **58**, 1599–1604 (1998).
48. Kriehuber, E. *et al.* Isolation and characterization of dermal lymphatic and blood endothelial cells reveal stable and functionally specialized cell lineages. *J. Exp. Med.* **194**, 797–808 (2001).
49. Karpanen, T. *et al.* Lymphangiogenic growth factor responsiveness is modulated by postnatal lymphatic vessel maturation. *Am. J. Pathol.* **169**, 708–718 (2006).
50. Saaristo, A. *et al.* Lymphangiogenic gene therapy with minimal blood vascular side effects. *J. Exp. Med.* **196**, 719–730 (2002).



# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

**A**t this year's Euroscience Open Forum meeting, held last week in Barcelona, Spain, a couple of the careers sessions saw panellists discuss unconventional career paths in science. The speakers included a patent lawyer, an entrepreneur, a scientist now working at a foundation and another at a government-led science initiative. Despite their divergent career tracks, they had quite a lot in common. Several spoke of the hard choices they faced when they committed to career changes both inside and outside the lab, and others emphasized the need for 'translators' to bridge the divide between fields or vocations.

The increasingly interdisciplinary approach to research has made scientists who can act as a bridge between two fields a hot commodity. But, as the Feature on page 662 reveals, forging a link between vocations can throw up some additional, complicating factors. Scientists who enter the political arena, in whatever capacity, must find ways to be effective science communicators while also realizing that science is only one of many elements that policy-makers need to take into account.

Interdisciplinary funding mechanisms — or in the case of politics and science, fellowships that spark in-depth dialogue between the two spheres — can help. But no funding mechanism can soften the culture shock that can come with joining and assimilating a different discipline.

The issue of hard choices was thrown into stark relief when one audience member asked the inevitable questions about balancing family and work life. Is it possible? How have the panellists accomplished this? Some less conventional science paths are more family friendly, noted one panellist, such as science publishing, which can offer regular hours and a manageable travel schedule. But a few panellists, female in particular, offered a rather bleak point of view. To excel in science, they said, the vast majority will have to choose to put their career before their family for at least some of their career. It's an excruciating choice to make, no matter what the discipline.

**Gene Russo is editor of *Naturejobs*.**

## CONTACTS

**Editor:** Gene Russo

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)  
**Business Development Manager:**  
Amelie Pequignot (4974)  
e-mail: [a.pequignot@nature.com](mailto:a.pequignot@nature.com)  
**Natureevents:**

Claudia Paulsen Young (+44 (0) 20 7014 4015)  
e-mail: [c.paulsenyoung@nature.com](mailto:c.paulsenyoung@nature.com)  
**France/Switzerland/Belgium:**  
Muriel Lestringuez (4994)  
**Southwest UK/RoW:** Nils Moeller (4953)

**Scandinavia/Spain/Portugal/Italy:**  
Evelina Rubio-Hakansson (4973)  
**Northeast UK/Ireland:**  
Matthew Ward (+44 (0) 20 7014 4059)  
**North Germany/The Netherlands:**  
Reya Silao (4970)  
**South Germany/Austria:**  
Hildi Rowland (+44 (0) 20 7014 4084)

**Advertising Production Manager:**  
Stephen Russell  
To send materials use London address above.  
Tel: +44 (0) 20 7843 4816  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)  
**Naturejobs web development:** Tom Hancock  
**Naturejobs online production:** Dennis Chu

**US Head Office, New York**  
75 Varick Street, 9th Floor,  
New York, NY 10013-1917  
Tel: +1 800 989 7718

Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:** Peter Bless

**India**  
Vikas Chawla (+91 1242881057)  
e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

**Japan Head Office, Tokyo**  
Chiyoda Building, 2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**  
Ayako Watanabe (+81 3 3267 8765)  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)  
**Business Development Manager, Greater China/Singapore:**  
Gloria To (+852 2811 7191)  
e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)

# Political will

Scientists who enter the world of political advocacy stand to gain perspective but could face a culture shock. **Gene Russo** reports.

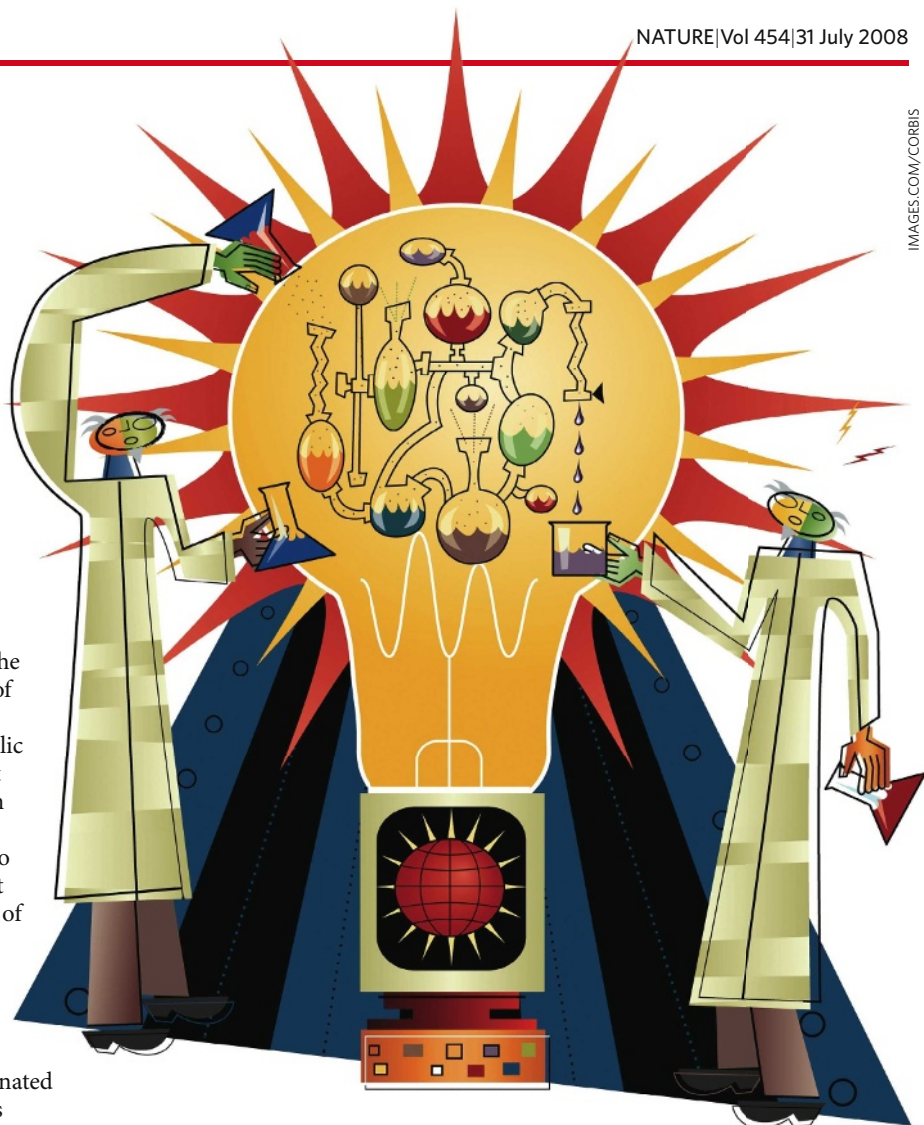
**W**hen climate scientist Dave Reay of the University of Edinburgh, UK, first appeared before a parliamentary select committee he was rather intimidated. He sat in a grand House of Commons room before a group of distinguished Members of Parliament (MPs), and a packed public gallery. Behind him sat one of his idols, naturalist David Attenborough. The room went quiet. Then came the first question from the chair: "What would your advice be, as the author of a book who has addressed this subject, in terms of how we get the citizens certainly of this country and, indeed, of other jurisdictions, involved in a practical way to address the climate change agenda?"

It was December 2006 and Reay was at the meeting to talk about what role everyday households could and should have in mitigating climate change. It was a sensitive subject that resonated with the public, and Reay had to choose his words carefully. He quickly learned the nuances of answering questions without going beyond the boundaries of his area of expertise (see "Tips for science advocacy").

## Clash of cultures

Whether driven by curiosity, self-interest or civic duty, many scientists like Reay dip into the world of advocacy and politics. Roles range from occasional letters in local newspapers to education sessions with policy-makers, to professional lobbying (to represent the interests of science institutions) or even elected office. Often these roles require the understanding of subtle principles of governance and policy-making. And often they mean a clash of cultures between science and politics — two disciplines whose modus operandi differ despite their many mutual interests. Those who enter the advocacy or political arena can expect frustration, occasional victories and a variety of challenges related to communicating science and understanding how policy-makers perceive it.

"There are risks involved in putting your head above the parapet, being seen as someone supporting a government policy or giving advice that others might disagree with," says Reay. President George W. Bush's rejection of the Kyoto Protocol is what pushed Reay to step in. He felt a responsibility to educate policy-makers about climate data, rising temperatures and the implications for the planet. He remains a full-time scientist and part-time advocate. The minor celebrity that came with his book *Climate Change Begins at Home* (Macmillan, 2006) has led to question-and-



answer sessions with select committees, and he has written synthesis papers for the UK and Scottish parliaments, including a summary of the main points of the Intergovernmental Panel on Climate Change's 4th assessment report.

Some UK scientists participate in an 'MP pairing' programme that matches a scientist and an MP so that both may learn from each other. The Parliamentary Office of Science and Technology takes PhD students for three-month stints to help inform the office's science-policy advice to Parliament.

## Delicate balance

Even in these seemingly innocent pursuits, scientists should tread carefully, say those familiar with the two worlds. Scientists are sometimes over-confident that scientific evidence will be the guiding principle in political decision-making. They may not fully appreciate the other economic and social factors affecting policy negotiations. And they may not wish to support a flawed initiative in order to win associated funds for more promising research.

Catherine Rudder, a professor of public policy at George Mason University in Arlington, Virginia, goes further. There is a real danger, she says, that scientists who step onto a political battlefield without understanding the issues and positions could be unwitting pawns in a game they would prefer not to play. Rudder, former executive director of the American Political Science Association, says the ends



Dave Reay has offered advice on climate change.



justify the means in politics. A scientist's authoritative, expert voice could be used by politicians to serve their own agenda.

This means that scientists providing testimony should be careful to not overstep their expertise. "You want to stay in the area where you're expert," says long-time science lobbyist Joel Widder, a member of Lewis-Burke Associates in Washington DC. "Stay in the science box." His firm's clients include universities such as the California Institute of Technology and the private stewards of national laboratories such as Fermilab's Fermi Research Alliance. He says that scientists should remain the "smartest people in the room" on the topics they engage in with policy-makers. They stand to lose credibility if they get into a debate on economic or policy issues they may not fully understand, he says.

A common tip for full-time scientists is to do some advocacy through their scientific society, which is already well-versed in science-policy machinations.

### Smartest in the room

Perhaps the biggest challenge is learning how to communicate complex scientific findings in clear, coherent and concise ways. Scientists who engage with political staff or legislators must know their audience, cautions Mike Holland, an analytical chemist turned policy analyst who has worked on the US House Committee on Science and Technology and the Office of Management and Budget. That audience could be as unsophisticated as a 20-something political-science graduate with little science education. Widder advises his scientist clients to pretend they are addressing readers of *The Wall Street Journal*: smart and well-read, but not necessarily familiar with the details of science.

In the United States, scientists have several options. Many societies sponsor congressional fellowships through the American Association for the Advancement of Science. These place scientists with representatives or senators, sometimes leading to a fully fledged policy or advocacy career. Interested scientists might also offer to help their university's government affairs office raise the institution's profile and vie for scarce funds on Capitol Hill.

Those wishing to dabble rather than devote an entire career might try the recently started Science, Health and Related Policies network founded by the non-profit Scientists and Engineers for America (SEA). The network has a web page for every member of Congress and senator, with their contact details and stances on science-related issues. The SEA has also started to form graduate student chapters at a few universities.

For scientists looking to have more influence and perhaps launch a political career, the advice seems simple: form partnerships and build relationships. Holland recommends that scientists should turn themselves into a resource for a local politician, explaining complex science issues. Young scientists can try volunteering with a campaign in their area, suggests physicist Michael Lubell, director of public affairs at the American Physical Society. His involvement in political campaigns stretches back to Lyndon B. Johnson's successful presidential campaign of 1964. Those who take part in winning campaigns may end up being reliable sources for advice on issues such as air quality or water treatment. European MP Malcolm Harbour says the European MP pairing programme — modelled

## TIPS FOR SCIENCE ADVOCACY

- Know your audience. Communicate your science in a clear, concise but intelligent manner.
- Consider other implications aside from just the budgetary — how should the science initiatives be prioritized?
- Recognize the perceptions of different fields and disciplines — for example, some US congressmen have a negative view of scientists associated with environmental groups.
- Be aware that explicit advocacy activities, especially if allied with a certain political party, could cause some tension with colleagues who disagree.
- Be careful when reaching outside of your area of expertise. Don't be afraid to state the limits of your knowledge on a subject.
- Consider advocating through a science society that knows the issues.
- Recognize that a full-time career move to advocacy could affect your prospects for returning to research.
- Recognize that other factors, such as values, jobs and economics, play into science policy. Laws rarely grow out of scientific evidence alone.

G.R.

on the UK version — has both opened participants' eyes to the complexities of governance, and opened his eyes to the complexities of managing massive science projects.

But to succeed at any level, researchers need to enjoy what can be a tedious process with many failures and false starts, says Geoff Mumford, director of science policy for the American Psychological Association. Research offers the satisfaction of generating peer-reviewed publications after years or months of hard work. But science advocates aiming for improved legislation or new initiatives may end up with nothing, time after time. "You have to enjoy the process and the strategic planning involved," Mumford says.

### Negative connotations

Any scientist who gets involved should be aware that the label 'advocate' or even 'political adviser' is not without its negative connotations. Brenda Ekwurzel studied climate variability as a professor at the University of Arizona in Tucson before deciding in 2004 to join the advocacy group the Union of Concerned Scientists, based in Cambridge, Massachusetts. Although she focuses on education more than advocacy, she concedes that returning to full-time research would be difficult, both because she has lost touch with the latest literature and because of the stigma attached to advocacy organizations. An academic position combining science, policy and politics might be a possibility, she says.

Lubell has managed to do some physics research while serving as an advocate, only recently taking a break for reasons unrelated to politics. "If you're a serious scientist, if you have a strong science reputation, I don't think anyone will accuse you of being affected by the politics," he says.

Reay believes his academic peers and superiors support his participation in politics, although he receives no official credit. But what if one's views differ radically from the science establishment's? "You could also destroy your career," he says with a chuckle. "It does depend on what you say."

**Gene Russo is editor of *Naturejobs*.**



**Taking care:** Brenda Ekwurzel (top) and Catherine Rudder.

H. FARRER



# MOVERS

**Joyce DeLeo, chair, Department of Pharmacology and Toxicology, Dartmouth Medical School, Hanover, New Hampshire**



**2003-present:** Professor of anaesthesiology and pharmacology, Dartmouth Medical School, Hanover, New Hampshire, USA

**2002-present:** Director, Neuroscience Center, Dartmouth Medical School

**1998-2003:** Associate professor of anaesthesiology and pharmacology, Dartmouth Medical School

Once a budding neuroanatomist, Joyce DeLeo made her mark by harnessing her interests in psychiatry and biology to search for novel pain treatments. As chair of Dartmouth Medical School's pharmacology and toxicology department, she now plans to use the same integrative approach to make progress in the department.

After a BS in biology and chemistry from the State University of New York in Albany, DeLeo was torn between graduate school and medical school. The independence she was given as a research assistant at Oral Roberts University in Tulsa, Oklahoma, helped her make up her mind.

DeLeo's dissertation, at the University of Oklahoma, focused on neuropharmacology related to stroke, and resulted in a patent for a drug to treat chronic pain. As the university's first Fulbright scholar, she next studied the electrophysiology of ischaemia at the Max Planck Institute for Psychiatry in Martinsried, Germany. The fellowship proved career-defining, as she studied the role of glial cells (the central nervous system's maintenance and support cells) in ischaemic stroke, under the leader of the field, departmental head Georg Kreutzberg.

A postdoc at Dartmouth Medical School's department of anaesthesiology began what has become a 20-year career so far. "Staying in one place doesn't limit you any more, technologically or collaboratively," she says. In fact, she says, it has afforded high-profile leadership opportunities.

DeLeo says she made her greatest progress once she began to read outside the pain field's specialist journals. She incorporated findings on the adaptive and innate immune systems into her search for novel agents to suppress the glial changes that produce chronic pain.

"Early on, Joyce was at the front of the pack, determining how the immune system might influence sensory perception and signalling in the spinal cord," says Michael Vasko, chair of the department of pharmacology and toxicology at Purdue University in Indianapolis, Indiana. He says DeLeo is up to the challenges of her new position, notably balancing basic versus translational research and bringing in funds.

"We want to find creative ways to fund our research, such as exploring new ways to partner with industry through patenting and licensing opportunities," DeLeo says. She plans to continue her pursuit of novel targets to treat and prevent pain by modulating glial function.

"It's important for chairs to lead by example — and she does that well in all areas," says Vasko.

**Virginia Gewin**

## NETWORKS & SUPPORT

### The gender imbalance

The US National Institutes of Health (NIH) this month launched a study into the root causes of gender disparity in scientific research. The study, which is being run by the National Institute of General Medical Sciences (NIGMS), has allocated between \$2 million and \$3 million to fund up to eight grants.

"At the lower level, women and men are about equal," says NIGMS director, Jeremy Berg. "But as you get farther and farther up the ladder, the number of women drop at each stage."

The issue of disparity is reasonably well documented, and previous studies have identified a number of reasons for this, such as women taking time off to have children. The NIGMS study aims to dig deeper to find out why there are differences between the career paths taken by men and women, says Berg. It will look at possible causes, such as family leave, institutional recruiting practices and the efficacy of programmes aimed at reducing gender discrimination.

The study follows in the wake of a report issued last year by the National Academy of Sciences. *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering* made a number of recommendations on how to increase

the participation of women in science. These included encouraging department heads at universities to be more accountable for their hiring policies, and for there to be campus programmes to help faculty members care for their children without stepping off the career ladder.

Berg notes that the NIGMS already has several programmes in place to help decrease ethnic disparity in sciences. Funding more research into gender disparity will help the agency see how it can adapt these approaches to women's issues, he says.

Janet Koster, executive director of the Association for Women in Science, has mixed feelings about the NIGMS study. She notes that the gender disparities have already been documented and that many organizations — including the National Academy of Sciences, the National Science Foundation and her own association — have issued recommendations to combat it. "It's great that the NIH has recognized the need to do something about the number of women who are leaving science," says Koster. "On the other hand, you can question the need for more research."

**Paul Smaglik**

**Join the debate at Nature Network**  
<http://tinyurl.com/65x54m>

#### POSTDOC JOURNAL

### 'Force'-full wisdom

"Do. Or do not. There is no 'try'." That's the advice Yoda gave Luke Skywalker in *The Empire Strikes Back*. This quote from the diminutive Jedi master sustains me as I head into an area beyond the expertise of my lab.

In addition to my ongoing work, I am now responsible for an unrelated mouse project. A postdoc friend initiated these studies, but due to various delays, he had to leave before the mice arrived. So I have inherited these animals — and the accompanying mountain of incomplete paperwork. The challenge is amplified by the fact that I've never worked with mice before; nor has anyone else in my lab.

I am immensely grateful for help from friendly mouse-geneticist neighbours, who have generously shared their expertise and time. I am excited to learn new skills, especially as biological research increasingly demands *in vivo* experiments as proof of physiological relevance. However, I am apprehensive about how this new project will turn out and how it will fit into my long-term career plan. Past experience with *Drosophila* and zebrafish projects has taught me that the knowledge and skills that I acquire are meaningless if no publication results from my work. So as I tackle yet another new model organism, I recall Yoda's words and fervently hope that the Force will be with me.

**Amanda Goh is a postdoctoral fellow in cell biology under the Agency of Science, Technology and Research in Singapore.**

# MOVERS

**Joyce DeLeo, chair, Department of Pharmacology and Toxicology, Dartmouth Medical School, Hanover, New Hampshire**



**2003-present:** Professor of anaesthesiology and pharmacology, Dartmouth Medical School, Hanover, New Hampshire, USA

**2002-present:** Director, Neuroscience Center, Dartmouth Medical School

**1998-2003:** Associate professor of anaesthesiology and pharmacology, Dartmouth Medical School

Once a budding neuroanatomist, Joyce DeLeo made her mark by harnessing her interests in psychiatry and biology to search for novel pain treatments. As chair of Dartmouth Medical School's pharmacology and toxicology department, she now plans to use the same integrative approach to make progress in the department.

After a BS in biology and chemistry from the State University of New York in Albany, DeLeo was torn between graduate school and medical school. The independence she was given as a research assistant at Oral Roberts University in Tulsa, Oklahoma, helped her make up her mind.

DeLeo's dissertation, at the University of Oklahoma, focused on neuropharmacology related to stroke, and resulted in a patent for a drug to treat chronic pain. As the university's first Fulbright scholar, she next studied the electrophysiology of ischaemia at the Max Planck Institute for Psychiatry in Martinsried, Germany. The fellowship proved career-defining, as she studied the role of glial cells (the central nervous system's maintenance and support cells) in ischaemic stroke, under the leader of the field, departmental head Georg Kreutzberg.

A postdoc at Dartmouth Medical School's department of anaesthesiology began what has become a 20-year career so far. "Staying in one place doesn't limit you any more, technologically or collaboratively," she says. In fact, she says, it has afforded high-profile leadership opportunities.

DeLeo says she made her greatest progress once she began to read outside the pain field's specialist journals. She incorporated findings on the adaptive and innate immune systems into her search for novel agents to suppress the glial changes that produce chronic pain.

"Early on, Joyce was at the front of the pack, determining how the immune system might influence sensory perception and signalling in the spinal cord," says Michael Vasko, chair of the department of pharmacology and toxicology at Purdue University in Indianapolis, Indiana. He says DeLeo is up to the challenges of her new position, notably balancing basic versus translational research and bringing in funds.

"We want to find creative ways to fund our research, such as exploring new ways to partner with industry through patenting and licensing opportunities," DeLeo says. She plans to continue her pursuit of novel targets to treat and prevent pain by modulating glial function.

"It's important for chairs to lead by example — and she does that well in all areas," says Vasko.

**Virginia Gewin**

## NETWORKS & SUPPORT

### The gender imbalance

The US National Institutes of Health (NIH) this month launched a study into the root causes of gender disparity in scientific research. The study, which is being run by the National Institute of General Medical Sciences (NIGMS), has allocated between \$2 million and \$3 million to fund up to eight grants.

"At the lower level, women and men are about equal," says NIGMS director, Jeremy Berg. "But as you get farther and farther up the ladder, the number of women drop at each stage."

The issue of disparity is reasonably well documented, and previous studies have identified a number of reasons for this, such as women taking time off to have children. The NIGMS study aims to dig deeper to find out why there are differences between the career paths taken by men and women, says Berg. It will look at possible causes, such as family leave, institutional recruiting practices and the efficacy of programmes aimed at reducing gender discrimination.

The study follows in the wake of a report issued last year by the National Academy of Sciences. *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering* made a number of recommendations on how to increase

the participation of women in science. These included encouraging department heads at universities to be more accountable for their hiring policies, and for there to be campus programmes to help faculty members care for their children without stepping off the career ladder.

Berg notes that the NIGMS already has several programmes in place to help decrease ethnic disparity in sciences. Funding more research into gender disparity will help the agency see how it can adapt these approaches to women's issues, he says.

Janet Koster, executive director of the Association for Women in Science, has mixed feelings about the NIGMS study. She notes that the gender disparities have already been documented and that many organizations — including the National Academy of Sciences, the National Science Foundation and her own association — have issued recommendations to combat it. "It's great that the NIH has recognized the need to do something about the number of women who are leaving science," says Koster. "On the other hand, you can question the need for more research."

**Paul Smaglik**

**Join the debate at Nature Network**  
<http://tinyurl.com/65x54m>

#### POSTDOC JOURNAL

### 'Force'-full wisdom

"Do. Or do not. There is no 'try'." That's the advice Yoda gave Luke Skywalker in *The Empire Strikes Back*. This quote from the diminutive Jedi master sustains me as I head into an area beyond the expertise of my lab.

In addition to my ongoing work, I am now responsible for an unrelated mouse project. A postdoc friend initiated these studies, but due to various delays, he had to leave before the mice arrived. So I have inherited these animals — and the accompanying mountain of incomplete paperwork. The challenge is amplified by the fact that I've never worked with mice before; nor has anyone else in my lab.

I am immensely grateful for help from friendly mouse-geneticist neighbours, who have generously shared their expertise and time. I am excited to learn new skills, especially as biological research increasingly demands *in vivo* experiments as proof of physiological relevance. However, I am apprehensive about how this new project will turn out and how it will fit into my long-term career plan. Past experience with *Drosophila* and zebrafish projects has taught me that the knowledge and skills that I acquire are meaningless if no publication results from my work. So as I tackle yet another new model organism, I recall Yoda's words and fervently hope that the Force will be with me.

**Amanda Goh is a postdoctoral fellow in cell biology under the Agency of Science, Technology and Research in Singapore.**

# MOVERS

**Joyce DeLeo, chair, Department of Pharmacology and Toxicology, Dartmouth Medical School, Hanover, New Hampshire**



**2003-present:** Professor of anaesthesiology and pharmacology, Dartmouth Medical School, Hanover, New Hampshire, USA

**2002-present:** Director, Neuroscience Center, Dartmouth Medical School

**1998-2003:** Associate professor of anaesthesiology and pharmacology, Dartmouth Medical School

Once a budding neuroanatomist, Joyce DeLeo made her mark by harnessing her interests in psychiatry and biology to search for novel pain treatments. As chair of Dartmouth Medical School's pharmacology and toxicology department, she now plans to use the same integrative approach to make progress in the department.

After a BS in biology and chemistry from the State University of New York in Albany, DeLeo was torn between graduate school and medical school. The independence she was given as a research assistant at Oral Roberts University in Tulsa, Oklahoma, helped her make up her mind.

DeLeo's dissertation, at the University of Oklahoma, focused on neuropharmacology related to stroke, and resulted in a patent for a drug to treat chronic pain. As the university's first Fulbright scholar, she next studied the electrophysiology of ischaemia at the Max Planck Institute for Psychiatry in Martinsried, Germany. The fellowship proved career-defining, as she studied the role of glial cells (the central nervous system's maintenance and support cells) in ischaemic stroke, under the leader of the field, departmental head Georg Kreutzberg.

A postdoc at Dartmouth Medical School's department of anaesthesiology began what has become a 20-year career so far. "Staying in one place doesn't limit you any more, technologically or collaboratively," she says. In fact, she says, it has afforded high-profile leadership opportunities.

DeLeo says she made her greatest progress once she began to read outside the pain field's specialist journals. She incorporated findings on the adaptive and innate immune systems into her search for novel agents to suppress the glial changes that produce chronic pain.

"Early on, Joyce was at the front of the pack, determining how the immune system might influence sensory perception and signalling in the spinal cord," says Michael Vasko, chair of the department of pharmacology and toxicology at Purdue University in Indianapolis, Indiana. He says DeLeo is up to the challenges of her new position, notably balancing basic versus translational research and bringing in funds.

"We want to find creative ways to fund our research, such as exploring new ways to partner with industry through patenting and licensing opportunities," DeLeo says. She plans to continue her pursuit of novel targets to treat and prevent pain by modulating glial function.

"It's important for chairs to lead by example — and she does that well in all areas," says Vasko.

**Virginia Gewin**

## NETWORKS & SUPPORT

### The gender imbalance

The US National Institutes of Health (NIH) this month launched a study into the root causes of gender disparity in scientific research. The study, which is being run by the National Institute of General Medical Sciences (NIGMS), has allocated between \$2 million and \$3 million to fund up to eight grants.

"At the lower level, women and men are about equal," says NIGMS director, Jeremy Berg. "But as you get farther and farther up the ladder, the number of women drop at each stage."

The issue of disparity is reasonably well documented, and previous studies have identified a number of reasons for this, such as women taking time off to have children. The NIGMS study aims to dig deeper to find out why there are differences between the career paths taken by men and women, says Berg. It will look at possible causes, such as family leave, institutional recruiting practices and the efficacy of programmes aimed at reducing gender discrimination.

The study follows in the wake of a report issued last year by the National Academy of Sciences. *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering* made a number of recommendations on how to increase

the participation of women in science. These included encouraging department heads at universities to be more accountable for their hiring policies, and for there to be campus programmes to help faculty members care for their children without stepping off the career ladder.

Berg notes that the NIGMS already has several programmes in place to help decrease ethnic disparity in sciences. Funding more research into gender disparity will help the agency see how it can adapt these approaches to women's issues, he says.

Janet Koster, executive director of the Association for Women in Science, has mixed feelings about the NIGMS study. She notes that the gender disparities have already been documented and that many organizations — including the National Academy of Sciences, the National Science Foundation and her own association — have issued recommendations to combat it. "It's great that the NIH has recognized the need to do something about the number of women who are leaving science," says Koster. "On the other hand, you can question the need for more research."

**Paul Smaglik**

**Join the debate at Nature Network**  
<http://tinyurl.com/65x54m>

#### POSTDOC JOURNAL

### 'Force'-full wisdom

"Do. Or do not. There is no 'try'." That's the advice Yoda gave Luke Skywalker in *The Empire Strikes Back*. This quote from the diminutive Jedi master sustains me as I head into an area beyond the expertise of my lab.

In addition to my ongoing work, I am now responsible for an unrelated mouse project. A postdoc friend initiated these studies, but due to various delays, he had to leave before the mice arrived. So I have inherited these animals — and the accompanying mountain of incomplete paperwork. The challenge is amplified by the fact that I've never worked with mice before; nor has anyone else in my lab.

I am immensely grateful for help from friendly mouse-geneticist neighbours, who have generously shared their expertise and time. I am excited to learn new skills, especially as biological research increasingly demands *in vivo* experiments as proof of physiological relevance. However, I am apprehensive about how this new project will turn out and how it will fit into my long-term career plan. Past experience with *Drosophila* and zebrafish projects has taught me that the knowledge and skills that I acquire are meaningless if no publication results from my work. So as I tackle yet another new model organism, I recall Yoda's words and fervently hope that the Force will be with me.

**Amanda Goh is a postdoctoral fellow in cell biology under the Agency of Science, Technology and Research in Singapore.**



# The pair-bond imperative

What's love got to do with it?

**Jennifer Rohn**

After the third time Eve came home late from work, her mother Mary activated the trace. On the small hand-held, she watched her daughter linger at the door of the domed bee farm, standing too close to a fair-haired young man. There was laughter; there was head-ducking; there was a hand stretched out to touch a forearm: her hand, his forearm. Five minutes of conversation, lips moving in animation. Then, after a display of reluctant body language, they parted like a broken comb of sticky honey.

For the next three days, Mary tuned in for clocking-off time. The man always arrived first and waited until Eve emerged, shy eyes brightening, from her designated station. They had known each other long enough to fall into a pattern, Mary saw. Despite the unexpected loveliness Eve radiated, this girl on the cusp of womanhood, her mother was both embarrassed and alarmed by the rawness of the emotions in evidence. For Mary did not share Eve's delight when she looked at the man. No, all she could see was that his sandy-blond hair was only a shade lighter than her daughter's bronzy head; streaked with red in the sun, yes, but still technically blond. And straight, equally fine and straight. They could have been brother and sister.

The colony had stopped DNA typing and other molecular techniques long ago, when food and energy production had become the main priority and such technology had been deemed an unsustainable luxury. Instead, they relied on the ancient methods practised by early scientists such as Gregor Mendel: phenotypic differences as a measure of genetic distinction.

After the catastrophic epidemic that had taken Eve's grandparents along with nine-tenths of the colony's inhabitants, the gene pool was too small to permit anything other than the most ruthlessly disparate matches; even then it was still unclear whether the group had a long-term future. The need to survive made mating choice another luxury. Women had to have children by as many different, and dissimilar, fathers as possible, for as many years as they were able. To allow pair-bonding

was impossible; attachments under such restrictions would be disastrous.

Eve knew the rules like everyone else, but, unlike her mother, when she looked at the man, who was called Paul, she stubbornly saw only differences. His hair was like nothing else on the planetoid: it was the hue of honeycombs held up to the sun, glowing coppery-gold, or perhaps the soft shade of citron fuzz on the insects that

had little conscious understanding of the underlying genetic imperative.

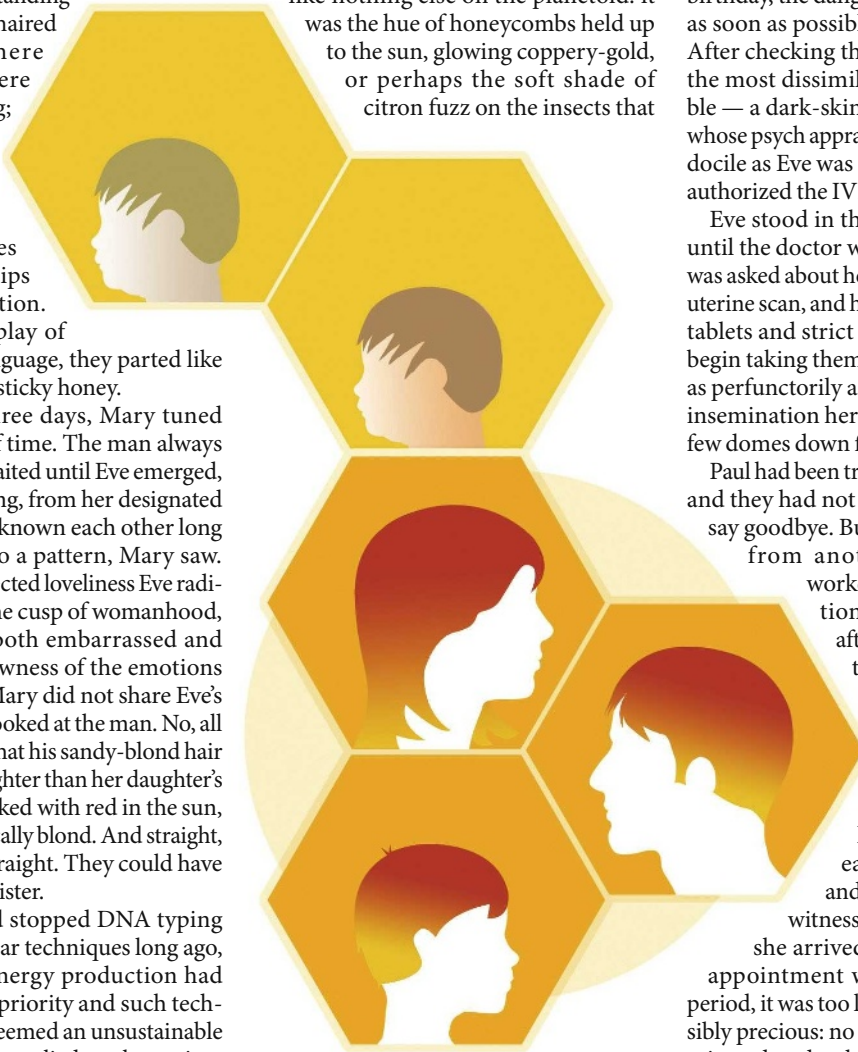
Meanwhile, Mary and her circle of half-sisters consulted with the local procreation officer. Yes, he agreed: although Eve was still a few months shy of her sixteenth birthday, the danger should be eliminated as soon as possible. Rules could be bent. After checking the database and finding the most dissimilar sperm donor possible — a dark-skinned, black-haired man whose psych appraisals suggested he was as docile as Eve was rebellious — the officer authorized the IVF personally.

Eve stood in the queue, close to tears, until the doctor was ready to see her. She was asked about her periods, given a rough uterine scan, and handed a vial of hormone tablets and strict instructions on how to begin taking them. And discharged, then, as perfunctorily as one of the sows whose insemination her half-brother oversaw a few domes down from the bee farm.

Paul had been transferred the day before, and they had not even had the chance to say goodbye. But Eve called in a favour from another half-sister who worked in the labour-allocation unit. Late that night, after flushing the hormone tablets down the toilet, she climbed out of her bedroom window and met him in the cornfields. The air was warm and redolent of newly ploughed earth; the moons had set and pale stars were the only witness. Four weeks later when she arrived at her egg-harvesting appointment with news of a missed period, it was too late. New life was impossibly precious: no pregnancy could be terminated under the rules of the colony, no matter how genetically imprudent.

The baby boy had hair halfway between bronze and copper, like silk on an ear of corn. And he gazed at Eve with eyes somewhere between the blue of midnight and dawn. She was never allowed to see Paul again, but somehow, with the child, this was bearable. And although Eve bore many more children, Mary secretly thought that this particular grandchild was the most beautiful and distinctive of all. ■

**Jennifer Rohn is a cell biologist at University College London and the editor of LabLit.com. Her first novel, *Experimental Heart*, comes out later this year.**



JACEY